PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

NOV 21 1928

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS

HAROLD RUGG, *Chairman.*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — L. M. TERMAN
FRANK N. FREEMAN — V. A. C. HENMON — BEARDSLEY RUML — E. L. THORNDIKE

VOL. XIX
No. 8

NOVEMBER, 1928

$5.00 a year.
70c per copy.

CONTENTS

THE EDUCATIONAL GROWTH OF ELEMENTARY SCHOOL CHILDREN. *Bird T. Baldwin* 521
THE MENTAL CAPACITY OF SOUTHERN ONTARIO INDIANS. *Elmer Jamieson and Peter Sandiford* 536
THE EFFECT UPON RETENTION OF VARYING LENGTHS OF STUDY PERIODS AND REST INTERVALS IN DISTRIBUTED LEARNING TIME. *Bernice E. Clark* 552
MATCHING ABILITY AS A FACTOR IN FIRST GRADE READING. *Nila Banton Smith* 560
RECIPROCAL RATINGS: A COMPARISON OF THE JUDGMENTS OF A TEACHER AND PUPILS IN ESTIMATING ABILITY IN UNITED STATES GEOGRAPHY. *Richard S. Uhrbrock* 572
EXPERIMENTAL COMPARISON OF TRUE-FALSE AND COMPLETION TESTS. *Violet Shulson and C. C. Crawford* 580
THEORIES UNDERLYING THE STATISTICAL DETERMINATION OF CREDIT TO BE ALLOTTED ITEM RESPONSES. *Elisha F. Bliss, Jr.* 584
THE SPEARMAN-BROWN PROPHECY FORMULA AND THE SEASHORE TESTS. *Paul R. Farnsworth* 586
NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES 589
NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION 593

Published Monthly Except June to August by

Warwick & York, Inc.,

York, Pa. Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at York, Pennsylvania under the Act of March 3, 1879.

The Journal of Educational Psychology

is published monthly, by Warwick & York, Inc., Baltimore, Md., and York, Pa. Title page and index are bound in December number.

Manuscripts for publication, books or other materials for review, and news items should be addressed to Harold Rugg, Chairman of the Editorial Board, 400 W. 118th Street, New York City.

The price of the Journal is $5.00 a year, payable in advance. Foreign postage is 30 cents extra. Single current issues are 70 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered. Back volumes $5.50 each; single back issues 90 cents each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which change is to take effect. No claim for nonreceipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European Plan

Entirely Fire-Proof

Centrally Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

The Commonwealth Teacher Training Study

By W. W. CHARTERS
and DOUGLAS WAPLES

Materials for Developing Teacher-Training Curricula and Methods of Supervision.

To be published in February
$3.00

THE
UNIVERSITY *of* CHICAGO PRESS
CHICAGO - - - ILLINOIS

The following institutions have adopted
CARTER V. GOOD'S

How To Do Research In Education

Published June 15, 1928

Bucknell Univ.
Columbia Univ. (Teachers College)
Cornell Univ.
Colorado State Teachers College
George Peabody College for Teachers
Harvard Univ.
Louisiana State Univ.
Loyola Univ.
Miami Univ.
New York Univ.
Northwestern Univ.
Ohio State Univ.
Temple Univ.
Univ. of California
Univ. of Chicago
Univ. of Cincinnati
Univ. of Denver
Univ. of Michigan
Univ. of Minnesota
Univ. of North Dakota
Univ. of Oregon
Univ. of Pennsylvania
Univ. of Pittsburgh
Univ. of Southern California
Univ. of Utah
Univ. of Washington
Univ. of Wichita
Yale Univ.

Price $2.50 + 10¢ post.

WARWICK AND YORK, INC.,
Publishers
BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XIX November, 1928 Number 8

THE EDUCATIONAL GROWTH OF ELEMENTARY SCHOOL CHILDREN[1]

BIRD T. BALDWIN[2]

Iowa Child Welfare Research Station, State University of Iowa

A comprehensive investigation of the relation of child development to school progress was begun by the writer in May, 1923, in the city of Cleveland. Five elementary schools were selected to represent various types of schools from the five districts of Cleveland. Children from the kindergarten through Grade VI were included; and each child was studied from many phases of his physical, mental, educational, and social development. The study had the direct cooperation of Superintendent Jones, Assistant Superintendent Bixby, and the other assistant superintendents, the supervisors, the heads of the different divisions of the school system, the principals, and the teachers.

The Aim, Scope, and Methods of Study

In this paper the aim is to summarize briefly only the educational growth of the pupils as determined by the application of the Stanford Achievement educational tests in these five schools for a three year period. Forms A and B of the primary and of the advanced Stanford Achievement tests were given to the children in the ten half-year grade divisions, IIB to VIA inclusive. In May, 1923, the test was given to 2445 children; in May, 1924 to 2557 children; and in May, 1925 to 2581 children, a total of 7583 educational achievement tests. On account of the tremendous turnover in these schools and the fact that only the children who were in the Grades II, III and IV during the

[1] A three-year study based on the Stanford Achievement test in the Cleveland Public Schools.

[2] The writer is particularly indebted to Carleton H. Mann, Helen V. Garside, and Naomi St. John for help in the preparation of this report.

first year of the study could be followed throughout, only 782 children (representing a total of 2346 tests) were tested for the three consecutive years. Of these, 244 were enrolled in Corlett school, 68 in Fruitland, 114 in Mill, 155 in Parkwood, and 201 in Washington Irving. All the 7583 tests are used in this paper except in the study of increments and of prediction, for which repeated tests are necessarily used.

Achievement by Schools and Grades

Composite Scores.—The educational status of the five schools may be determined by a comparison, by grade divisions, of the scores on the Stanford Achievement test with the norms on the test published in April, 1924, and by a comparison with each other for the three consecutive years. The norms of 1927 differ from those of 1924 for the age groups represented in this study by an average deviation of 0.6 month. In the first year, the medians of the total composite scores of the five schools in the lower grades are above the test norms, but they become increasingly closer in the upper grades, with slight deviations, until grade division VIA, when they drop below. Parkwood school has the highest median in each half-year grade division with a range from 4.7 points above the grade norms in Grade IIIA to 11.5 points in Grade VIA. Corlett is lowest in all the half-year grade divisions except IVB, VIB and VA. In the second year the status of the schools remains the same. In the third year, Fruitland and Parkwood are above the test norms in all grades.

A comparison of the median composite scores by grade divisions for the three years with the test norms shows that the composite scores for these years are higher than the test norms in all grades except in grade division VIA in the first year, but they tend to approach the test norms in the upper grade divisions (Table I).

The high percentage of foreign-born population in some of these schools doubtless is the chief cause for their relatively lower educational standing, and should not reflect on the effectiveness of the work done by the schools. At the beginning of this study there were approximately eighty per cent foreign-born in Corlett, forty-four per cent in Washington Irving, thirty-three per cent in Mill, and thirty per cent each in Fruitland and Parkwood. About two-thirds of the foreign-born in Parkwood were Russian Jews. In addition to the forty-four per cent foreign-born in Washington Irving, twenty-six per cent were negroes.

TABLE I.—MEDIAN COMPOSITE SCORES BY SCHOOLS AND GRADE DIVISIONS ON THE STANFORD ACHIEVEMENT TEST, FORMS A AND B, PRIMARY AND ADVANCED EXAMINATIONS

Grade division	Corlett		Fruitland		Washington Irving		Mill		Parkwood		All schools				Test norm
												Scores			
	Number of children	Median score	Number of children	Median score	Number of children	Median score	Number of children	Median score	Number of children	Median score	Number of children	Twenty-five percentile	Median	Seventy-five percentile	
Primary and Advanced Examinations, 1923, Form A															
2B	63	7.6	18	9.8	79	8.6	36	12.0	42	15.1	238	7.1	9.8	13.5	5.0
2A	99	14.4	21	16.8	95	16.1	30	17.5	49	18.3	294	12.8	16.2	19.2	12.0
3B	54	16.4	19	21.5	76	23.5	36	21.0	60	25.9	245	17.1	21.5	26.4	19.0
3A	84	22.8	20	27.0	91	25.8	40	23.5	31	28.8	266	20.1	25.3	28.9	24.0
4B	61	37.4	18	39.0	70	31.6	40	33.0	43	42.1	232	29.6	35.5	41.7	32.0
4A	90	36.0	33	46.3	75	37.1	33	38.8	50	47.0	281	34.2	38.9	45.4	38.0
5B	67	42.2	23	48.8	78	43.5	24	42.0	38	50.7	230	39.4	44.9	50.5	43.0
5A	56	48.3	19	56.2	75	47.0	26	50.0	65	58.3	241	46.3	51.0	56.6	49.0
6B	51	51.2	17	58.1	42	54.5	31	57.3	35	64.1	176	49.9	55.7	61.2	54.0
6A	81	54.2	13	57.5	75	55.5	36	60.3	37	69.5	242	51.2	57.4	65.5	58.0
Primary and Advanced Examinations, 1924, Form B															
2B	65	8.8	21	10.4	67	9.7	28	13.5	38	12.8	219	7.4	10.5	14.0	5.0
2A	67	12.7	24	14.7	78	15.3	31	15.9	62	20.1	262	11.9	15.7	19.4	12.0
3B	83	17.6	22	23.0	67	20.6	33	16.7	47	24.5	252	16.6	20.1	23.4	19.0
3A	86	21.6	36	27.0	73	24.9	31	22.5	58	28.0	284	20.7	24.8	29.1	24.0
4B	69	31.1	13	31.5	75	33.8	45	36.2	62	41.0	264	28.8	34.1	40.4	32.0
4A	95	34.9	26	44.0	98	40.0	36	37.5	42	41.5	297	33.7	38.5	43.6	38.0
5B	68	38.0	19	49.2	52	44.5	36	42.0	44	47.6	219	37.7	44.6	49.9	43.0
5A	76	46.3	34	49.5	62	47.7	30	55.3	57	56.1	269	44.0	50.0	56.0	49.0
6B	63	53.2	23	58.7	71	55.9	27	55.1	46	60.7	230	49.1	56.2	62.1	54.0
6A	61	59.2	15	66.2	81	56.7	30	60.0	74	67.5	261	54.8	61.9	67.8	58.0
Primary and Advanced Examinations, 1925, Form A															
2B	71	11.3	24	9.5	60	8.3	42	9.0	40	10.8	237	7.0	10.0	13.5	5.0
2A	90	15.0	32	15.7	67	17.6	31	13.5	65	19.8	285	12.8	16.8	20.5	12.0
3B	64	17.0	23	21.4	66	21.7	39	22.2	40	25.7	232	17.0	21.7	25.4	19.0
3A	72	22.6	25	27.8	68	25.7	41	26.4	65	29.9	271	22.3	26.2	29.8	24.0
4B	84	33.3	17	36.5	88	33.4	35	35.4	48	40.5	272	29.3	34.7	40.0	32.0
4A	87	37.8	41	42.5	71	37.6	36	38.5	51	47.5	286	34.7	39.8	46.1	38.0
5B	66	40.6	17	47.5	62	41.7	41	40.8	74	51.5	260	38.3	44.4	51.6	43.0
5A	84	51.1	24	53.5	82	46.8	35	46.1	46	54.0	271	44.0	49.6	55.2	49.0
6B	71	49.8	16	57.5	46	53.7	29	56.3	46	59.3	208	49.0	54.5	61.1	54.0
6A	84	56.9	25	60.5	43	60.2	36	65.0	71	67.1	259	55.8	61.8	67.8	58.0

From the data presented the following deductions are drawn: (1) There is a marked difference in achievement in schools in this group; (2) on the whole the schools and grades maintain their relative standing for the three years; (3) Parkwood excels; (4) Corlett ranks lowest; (5) the average of the grade medians for the five schools exceeds the test norms at all points; (6) median scores by grades exceed in general the test norms by a greater number of points in the lower grades than in the upper grades; and (7) median scores by grades for the three years approximate each other very closely.

Subject Scores.—On the median and percentile scores in reading, arithmetic, and spelling, subjects that are included in both the primary and advanced forms of the Stanford Achievement test, the five schools have the same relative ranking as on composite scores; this fact shows that there is no particular subject weakness in any one school, but that the general educational standard of each school is fairly stable and is maintained in all subjects.

Achievement Level by Schools and Age

The high percentage of over-age pupils in the Cleveland schools,[1] due chiefly to the large foreign-born population, is undoubtedly a very important factor in causing the lower scores for the upper ages; and since the foreign-born over-age pupils probably have language and other difficulties, their large number may also be important in causing the lower scores for the upper grades. In 1922–1923, over-age pupils totaled thirty-four per cent, at age, 60.4 per cent, and under-age 5.6 per cent. With approximately one-third of the pupils over-age, it is obvious that the schools are adapting their courses of study and methods of promotion to the needs of these pupils. It is evident that chronological age alone is not a significant factor in predicting attainment. It is clear that pupils of a chronological age characteristic of their grade are farther advanced in attainment than pupils a year younger and that in any grade the pupils who are older than the age characteristic of their grade are there primarily because of their low attainments. The statement may be made, then, that in general in the primary examination, for the younger ages there is a fairly steady increase in score with age, and in the older ages a steady decrease in

[1] Over-age was ascertained by the Strayer scale. Strayer, George: Age and Grade Census of Schools and Colleges. *United States Bureau of Education Bulletin* 5, 1911, p. 12.

score with age, the over-age pupils failing to meet the standard for the norm.

In the advanced examination the same general trend occurs. Up to the age group 11–6 to 11–11, when the first group is omitted, there is an unbroken increase in score. From this point on the trend is downward with slight rises at age groups 13–6 to 13–11 and 14–6 to 14–11. If all the age groups are separated into two groups at twelve years, the average of the medians of the age groups before twelve in the advanced examination is 50.0 points, and of the groups after twelve, 49.1.

Distribution of Scores by Grades

When the composite scores for the ten grade divisions are grouped in steps of five points each, there is a wide distribution of scores for

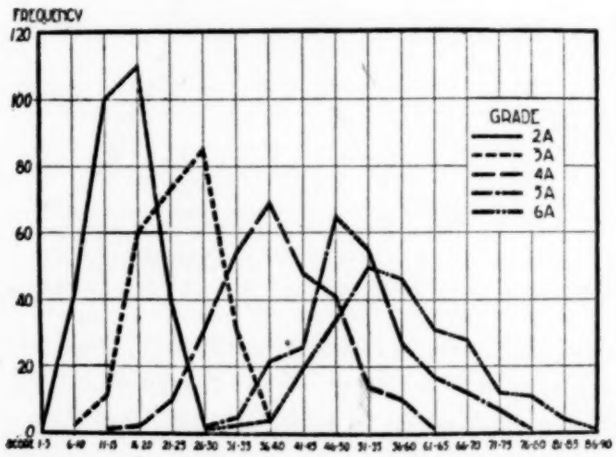

Fig. 1.—The distribution of composite scores by grades on the Stanford Achievement test shows marked overlapping.

each grade, and the range becomes gradually wider in the upper grades. Figure 1 illustrates the marked overlapping of scores. For example, in the 31–35 score group four grades, including seven half-year grade divisions, are represented. In the 26–30 score group all grade divisions from IIA to VIA are represented. Further analysis of the overlapping reveals the fact that in grade division IVA, 23.4 per cent of the children are doing as well as or better than 24.3 per cent in VIA. These results are in general accord with the overlapping of achievement of children from one grade to another and one division to another and show that it is impossible to grade a school on any one particular type of achievement or phase of development. However, there are some

children in grade division IVA who, no doubt, could be promoted to VA or VIA to advantage.

Yearly Increments in Composite Scores

From an educational point of view, the value of the application of the Stanford Achievement test lies not only in determining the educational status of the schools as shown by a cross section of the results for

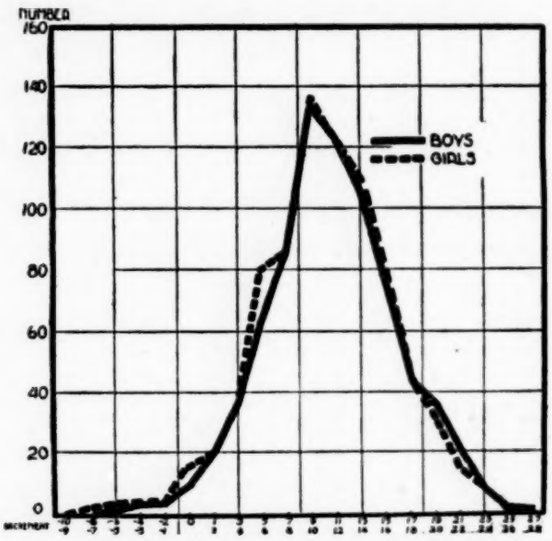

Fig. 2.—The increments on composite scores for boys and for girls during the two-year period 1923 to 1925 reveal very slight sex differences.

a particular year, but also in determining specifically the yearly increments of composite scores by age and sex for each school, for the schools as a whole, and for each child in the schools. In this study the widest range in increment for the total group during the first yearly increment period is in the 9–10 age group, extending from the −9 to −10 score group to the +23 to +24 score group, a range of thirty-four points; during the second yearly increment period the widest range is in the same score group, and slightly wider.

The sex differences are insignificant; the girls show slightly more irregularity in increments than the boys (see Fig. 2). The girls show a total irregularity, obtained by adding the variations in increment by age groups between the two periods, of 40.8 points on median increments and the boys a total irregularity of 36.1.

The median increments by age and sex for the second and third years are given in Table II. A comparison of the median increments for the schools and the relative educational growth of the schools for two periods, 1923 to 1924 and 1924 to 1925, is given in Fig. 3.

TABLE II.—MEDIAN INCREMENTS ON COMPOSITE SCORES ON THREE CONSECUTIVE STANFORD ACHIEVEMENT TESTS BY AGE AND SEX

Age in 1923, years	Number	Increment			
		1923–1924, second test over first	Normal for age	1924–1925, third test over second	Normal for age
Boys					
6	2	8.0	8.5	15.5	7.0
7	71	10.9	7.0	15.5	8.0
8	112	12.7	8.0	12.1	9.0
9	111	11.6	9.0	10.5	12.0
10	57	10.5	12.0	10.4	11.0
11	23	9.8	11.0	10.5	11.0
12	6	2.0	11.0	11.5	7.0
13	1	8.5	7.0	13.5	5.0
14	2	10.5	5.0	3.5	2.5
All ages	385	11.5	8.7	11.6	8.1
Girls					
6	1	5.5	8.5	18.5	7.0
7	89	9.6	7.0	15.3	8.0
8	121	13.4	8.0	10.8	9.0
9	113	11.0	9.0	11.1	12.0
10	57	7.8	12.0	9.9	11.0
11	11	8.5	11.0	11.8	11.0
12	1	−2.0	11.0	4.5	7.0
13	2	6.5	7.0	7.0	5.0
14	2	6.0	5.0	13.0	2.5
All ages	397	10.8	8.7	11.6	8.1

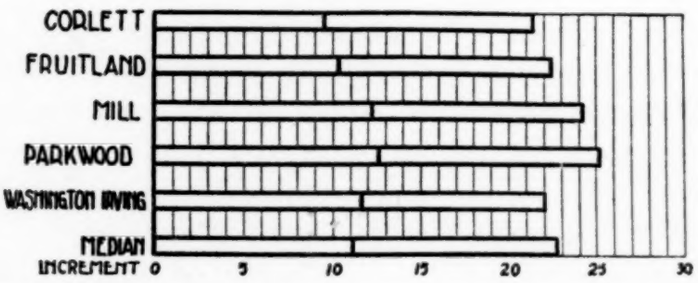

FIG. 3.—The median educational growth for each of the five schools, based on the median yearly increments on composite scores, is compared with the median increments for all five schools.

Influence of Practice and Instruction Revealed through Retests

The mean scores for children given consecutive tests shows the effect of practice and any instruction given by the teachers that may have affected the content of these tests (Table III). For example, the average score for boys at age seven at the time of the first test in 1923 is 11.6; at the time of the first re-test in 1924 when these boys were eight years old, it is 22.4; while the average for all boys of age eight at the time of the first test is only 16.0. Similar increases are made by

Table III.—Means and Standard Deviations of Scores on Three Consecutive Stanford Achievement Tests

Age in 1923, years	Number	1923		1924		1925	
		Mean score	SD	Mean score	SD	Mean score	SD
Boys							
7	25	11.6	5.04	22.4	8.54	38.2	9.93
8	100	16.0	6.75	27.4	10.13	41.0	11.92
9	117	25.1	10.20	37.9	11.58	49.0	12.18
10	96	33.1	11.58	44.2	12.88	54.2	13.28
11	33	33.4	7.88	44.3	8.71	53.2	9.26
Girls							
7	30	12.4	5.94	21.5	7.57	37.3	8.93
8	115	18.3	7.80	29.1	10.71	42.4	10.70
9	126	26.1	9.94	38.3	10.48	49.1	11.32
10	87	35.0	10.95	44.2	10.98	54.4	12.78
11	26	33.8	7.96	42.9	10.11	53.2	11.84

all the pupils on the second and third tests; consequently the mean age score was considerably raised for each of the three years for the five schools, and the average score increased above the expected gain for each age for the same group of pupils.

Correlations of Educational Scores for Three Consecutive Years

Significance of the Correlations.—The high correlations and low probable errors given in Table IV indicate that variation due to chance

sampling is low in this study of the relation of Stanford Achievement scores of a group at a given year and given age and the scores of the

TABLE IV.—CORRELATIONS ON THREE CONSECUTIVE STANFORD ACHIEVEMENT TESTS

Age in 1923, years	Number	1923–1924		1923–1925		1924–1925	
		r	PE	*r*	PE	*r*	PE
Boys							
7	25	.84	±.04	.86	±.03	.94	±.02
8	100	.92	±.01	.91	±.01	.90	±.01
9	117	.90	±.01	.86	±.02	.91	±.01
10	96	.91	±.01	.90	±.01	.94	±.01
11	33	.78	±.04	.80	±.04	.92	±.02
Girls							
7	30	.87	±.03	.88	±.03	.82	±.04
8	115	.91	±.01	.85	±.02	.88	±.01
9	126	.85	±.02	.85	±.02	.91	±.01
10	87	.92	±.01	.88	±.02	.97	±.00
11	26	.82	±.04	.74	±.06	.89	±.03

same group one or two years later. The slightly larger probable errors at ages seven and eleven years show that at these ages the coefficients of correlation are subject to greater fluctuation than at the other ages. This is probably due to the smaller number of cases rather than to any real age factor. The chances are even that the coefficient of correlation, +.92, between the scores the eight year old boys made in 1923 and the scores they made when retested in 1924 would not fluctuate more than ±.01 if a similar sample were taken, and that if an indefinitely large number of similar samples were taken, fifty per cent of the correlations based on these samples would fall between .91 and .93.

In this study the coefficients of correlation are many times their probable errors and so indicate a high degree of association between the score made in a given year and the score made one or two years later.

As criteria of reliability of the test these correlations are a little low. Since the scores on the second series of tests contain the original measures or amounts of achievement as well as the gains, the correlations are too large to give a true idea of the relation between the gains and the original scores. Even were there no relation whatever

between the gains and the original measures, the correlation between the two series of measures would still be theoretically +.707, indicating a relation of fifty per cent between the two series of scores where no relation existed between the original scores and gains. The correlations of about .90 indicate an increase in the relationship of only about seventeen per cent. Corrected for attenuation, the relation between the original scores and gains would run up to only about twenty per cent.

Prediction of Educational Achievement

Prediction from Scores for One Year.—The achievement scores which boys and girls of various age groups from seven to eleven made in 1923 were correlated with the scores they made one and two years later. Judging the possible prediction of scores one or two years later in the light of the standard error of estimate, $S_y = \sigma_y(1 - r^2_{xy})^{\frac{1}{2}}$, it was found that the chances were two to one that the estimated scores for eight-year old boys, for example, would not deviate more than ± 4.0 points from the observed scores when 1924 scores were predicted from 1923 scores and by not more than ± 5.0 points from the observed scores when 1925 scores were predicted from 1923 scores.

Table V.—Standard Error of Estimate for Stanford Achievement Test Scores

Age in years, 1923	S_y	
	1924 from 1923 score	1925 from 1923 score
Boys		
7	4.6	5.1
8	4.0	5.0
9	5.0	6.2
10	5.3	5.8
11	5.4	5.6
Girls		
7	3.7	4.2
8	4.4	5.6
9	5.5	6.0
10	4.3	6.1
11	5.8	8.0

It is evident from Table V that the standard error of estimate is slightly larger for estimated scores for 1925 on the basis of scores made in 1923 than for estimated scores for 1924. But considering the standard error of estimates when they are weighted with the reciprocals of their respective mean scores, the estimation of 1925 scores from 1923 scores seems just as good as the estimation of 1924 scores from 1923 scores; that is, differences in length of time between tests do not seem to affect significantly the standard error of estimation.

For verification of this theoretical prediction, the scores for the eight-year group of boys were predicted for 1924 and for 1925, by use of the following regression equation in which the known variable was the 1923 scores.

$$Y - \bar{Y} = r\frac{\sigma y}{\sigma x}(X - \bar{X})$$

When the numerical equivalents were substituted, the equation used for the 1924 scores was $Y = 1.38 \times +5.38$ and for the 1925 scores $Y = 1.61 \times +15.37$. It was found empirically that seventy-four per cent of the estimated 1924 scores from 1923 do not deviate more than ± 4.0 (one standard error) from the observed scores, and seventy per cent of the estimated 1925 scores from 1923 scores do not deviate more than ± 5.0 (one standard error) from the observed scores.

Prediction from Scores for Two Years.—The 1925 scores for the eight-year group of boys were predicted also by use of the multiple regression equation in which both 1923 and 1924 scores were the known variables:

Letting the subscript $_0 = 1925$
$_1 = 1924$
$_2 = 1923$

$$\bar{X}_0 = B_{01.2}\frac{\sigma_0}{\sigma_1}X_1 + B_{02.1}\frac{\sigma_0}{\sigma_2}X_2 + (M_0 - B_{01.2}\frac{\sigma_0}{\sigma_1}M_1 - B_{02.1}\frac{\sigma_0}{\sigma_2}M_2)^*$$

When the numerical equivalents were substituted, the regression equation was $\bar{X}_0 = .479X_1 + .941X_2 + 12.87$. In this prediction sixty-three per cent of the cases do not deviate from the observed scores more than ± 4.5, one standard error ($S_{0.12}$) when $S_{0.12} = \sqrt{\sigma_0^2(1 - r_{01}^2)(1 - r_{02.1})}$.

There is a difference of only 0.5 point between the standard error in estimating 1925 scores from 1923 scores alone and in estimating 1925 scores from both 1923 and 1924 scores. The fact that this increase in

* Kelley; "Statistical Method," p. 283.

predictability is so slight may be accounted for by the small degree of relationship between the original scores and gains.

For some specific cases it is interesting to note the difference in the estimation when based on the knowledge of one score and when based on the knowledge of two scores. For example, if a child is advancing true to form, his future scores may be determined from the first rating almost as accurately as from the first and second scores combined.

The predicted score for child 324 for 1924 from his score for 1923 differs only by 0.6 point from the observed score. The predicted score for 1925 from the score for 1923 and the observed score are the same. The predicted score for 1925 from both 1923 and 1924 scores differs by 0.3 point from the observed score. For this child one measurement seems to be sufficient to make the predictions for the later years.

TABLE VI.—OBSERVED AND PREDICTED STANFORD ACHIEVEMENT TEST SCORES FOR TEN BOYS EIGHT YEARS OF AGE IN 1923

Boy's serial number	Age	1923 score observed,	1924 scores		1925 scores		
	Years, months		Observed	Predicted from 1923 score	Observed	Predicted	
						From 1923 score	From 1923 and 1924 scores
272	7- 6	14.6	16.9	25.5	36.2	38.9	34.7
115	7- 7	5.8	12.5	13.4	22.5	24.7	24.3
250	7- 7	16.5	24.2	28.2	46.3	41.9	40.0
252	7- 9	19.0	23.8	12.7	46.0	46.0	42.1
204	7-10	13.0	23.4	23.3	42.5	36.3	36.3
324	7-10	12.6	22.2	22.8	35.7	35.7	35.4
563	8- 0	11.6	14.1	21.4	21.7	34.0	30.5
64	8- 1	9.1	21.3	17.9	32.9	30.0	31.6
5	8- 2	25.9	40.5	41.1	57.0	57.1	56.6
403	8- 4	12.6	30.6	22.8	33.1	35.7	39.4

For child 204 the predicted score for 1924 from the score for 1923 differs only by 0.1 point from the observed score. The predicted score for 1925 from the score for 1923 and from both 1923 and 1924 is 36.3. The observed score, however, is 42.5, an unexpected increase, on the basis of the child's previous record.

Individual Profiles

When an intensive study of individual pupils is made and their scores expressed graphically in profiles indicating the relation of chronological to educational age, a clear insight is gained into the achievement of the pupils in the various subjects, their relation to the norms for the various subjects, and their annual increment in educational growth. The profile in Fig. 4 is given as a sample of the individual records upon which this study is based. The child from whose record this profile is made is accelerated in educational age. She made a larger relative gain by year than the children whose educational and

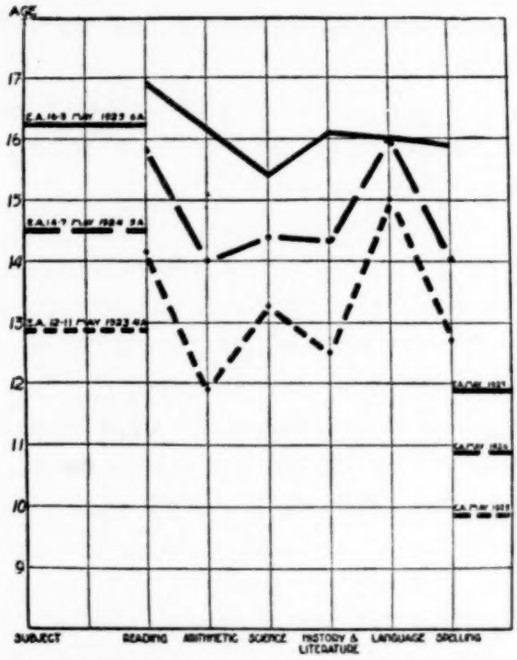

Fig. 4.—Consecutive individual profiles in educational growth in five subjects show the variations for three years for an accelerated pupil.

chronological ages are more nearly equal. In the twenty-four month period covered by this study she gained three years and four months educationally, an average educational increase of 1.7 months for each calendar month. She shows a very even increase in all subjects except in language for the last year.

Summary and Conclusions

In a cross section of the achievement of five Cleveland schools for three years based on results on the Stanford Achievement test for approximately two thousand five-hundred children each year, in grade divisions IIA to VIA, it is found that:

1. There is an appreciable difference in median composite achievement scores by grade divisions for the five schools.

2. On the whole, the schools and grade divisions maintain their relative standing in median composite scores throughout the three years.

3. When the median composite scores for the five schools combined are compared by grades with the test norms, it is found that the median scores are higher than the test norms. The median composite scores for the five schools gradually approach the test norms more closely as they ascend from grade divisions IIA to VIA.

4. The relative standing of the median composite scores for the five schools by age remains somewhat the same as the standing of scores by grades.

5. A marked overlapping of scores is found in all grades. The extreme example is in score group 26–30, in which all grade divisions from IIA to VIA are represented.

In a cross section of the achievement of the five Cleveland schools, based on three consecutive records on the Stanford Achievement test for seven hundred eighty-two children, these facts are found:

1. There is a wide distribution of yearly increment of scores.

2. The median increments on composite scores for boys and girls in the three years show insignificant sex differences.

3. The median increments on composite scores for 1924–1925 are slightly higher than for 1923–1924. This shows that the pupils have profited by the previous year's testing, since the mean age score is also raised each year by consecutive tests.

4. High correlations and low probable errors based on consecutive scores indicate that variation due to chance sampling in this study is low.

5. High correlations and low probable errors indicate a high degree of association between a score made by a pupil in a given year and a score made one year or two years later.

6. The standard errors of estimating 1924 scores from 1923 scores range for boys from ± 4.0 at eight years to ± 5.4 at eleven years, and for girls from ± 3.7 at seven years to ± 5.8 at eleven years. These standard errors of estimate are well within one year's educational growth. Empirically, seventy-four per cent of the estimated 1924 scores do not deviate more than ± 4.0 from the observed 1924 scores for eight-year old boys.

7. The length of time between tests does not seem to affect significantly the standard error of prediction, since for the eight-year old boys, it is found empirically that seventy per cent of the estimated 1925 scores from 1923 scores do not deviate more than ± 5.0 from the observed 1925 scores.

8. The individual educational profiles give a good diagnosis of the subject levels of achievement for each pupil, and indicate that there is a greater increment for pupils whose educational age is higher than their chronological age than for pupils whose educational age and chronological age are more nearly the same.

The facts revealed in this paper justify the statement that a study of the results of the Stanford Achievement test given for a series of years to the same pupils from Grades II to VI inclusive furnishes a good insight into their educational growth.

THE MENTAL CAPACITY OF SOUTHERN ONTARIO INDIANS[1]

ELMER JAMIESON AND PETER SANDIFORD

University of Toronto, Toronto, Canada

The discovery of economical and reliable methods of measuring the intelligence and the educational achievement of pupils has placed new weapons in the hands of those who wish to compare the intelligence and achievement of different races. Up to the present, however, few studies of the Indian have been made.

Rowe ('14), using the Goddard Revision of the Binet Scale, measured the intelligence of two hundred sixty-eight children in the state of Michigan. He found that only 5.8 per cent of the Indian children tested at or above age, compared with seventy-nine per cent for white children. It is probable that the language handicap was unduly great for Indian children.

Garth ('25), assisted by Serafini and Dutton, tested one thousand fifty full blood Indian children in the United States Indian Schools of Oklahoma and New Mexico, employing the National Intelligence Test, Scale A, Form 1. The median IQ of the United States Indians was found to be 68.6, and only 1.6 per cent of them were found to equal or exceed the median IQ of the whites. As this same test was used in the Ontario study an interesting comparison between American and Canadian Indians can be made.

Fitzgerald and Ludeman ('26) tested forty-one Indian children in a mission school in South Dakota and forty-two high school Indian children in a training school of Nebraska. They used the National Intelligence and the Terman tests at the mission school and the Otis test at the training school. The median IQ of 87.5 which was found, is high for Indian children and points to some sort of selection of superior Indian children in these schools.

To ascertain whether or not there are differences in the intelligence of Indians of different blood stocks, Garth ('23) administered the

---

[1] This summary was written by Sandiford. The investigation was carried out by Jamieson under Sandiford's direction. In the near future Jamieson will make the detailed results available in an extended printed report. It should be mentioned that Jamieson is an Indian who has resided on the Six Nations Reserve from birth and that he can speak the Mohawk dialect. He is a Master of Arts of McMaster University.

National Intelligence Test, Scale A, to one hundred twenty-six Mixed Bloods, three hundred seven Mexicans, two hundred forty-nine Plains and Forest Indians, one hundred seventy-six Pueblo Indians, and eighty-five Navajo and Apache Indians, the last three groups being full bloods. According to average and median scores as well as percentiles the five groups ranked in the following order, the most intelligent being placed first: (1) Mixed blood; (2) Mexicans; (3) Plains and Forest Indians; (4) Pueblos; (5) Navajo and Apache. The full-blooded Indians, who use more Indian than English, were probably handicapped by the nature of the tests.

Hunter and Sommermier ('21), using the 1911 edition of the Otis test with seven hundred fifteen Indians at Haskell Institute, Lawrence, Kansas, report that the median score for the Indians is 82.64 with a standard deviation of 36.7, while the median score of one thousand three hundred sixty-six fifteen-year white children is 122.58, with a standard deviation of 30.9. The quartiles for total score, the median score on each of the ten tests making up the scale, and the percentage of individuals testing at and above the Otis age norms, are found to decrease as the amount of Indian blood increases.

Garth ('22) in another of his many studies of the mental capacity and characteristics of Indian children, formulated the problem: "If groups of mixed and full blood Indians with the same school training as indicated by grade attainment are given the same psychological tests under controlled conditions, what will be the results of a comparison of their performances or the tests?" He sought to answer the question by administering four association tests, three memory tests, and two word-building tests to Indian pupils in the United States Indian school at Chiloco, Oklahoma. He reports that the mixed blood Indians excel the full bloods by about eleven per cent, but that there is considerable overlapping, and that the mixed bloods are favored by superior social status.

In the present study the subjects investigated were Indian pupils, all of whom were residents of Southern Ontario except eleven Caughnawaga Indians living near Montreal in the Province of Quebec. They were in attendance in the eleven day schools of the Six Nations, in the Mohawk Institute near Brantford, and in the Mount Elgin Residential School near St. Thomas. The schools of the Six Nations are undenominational day schools—typical rural schools of an Ontario farming community. Both the Mohawk Institute and the Mount Elgin School are residential farm schools, accommodating in preference orphans,

children of poor parents and other unfortunates. In addition to a general education the boys of the residential schools receive a practical training in farming and the girls a training in domestic economy.

The inhabitants of the Six Nations Reserve are Iroquois (a confederacy of six tribes, namely, Mohawks, Cayugas, Oneidas, Onondagas, Senecas and Tuscaroras). The Delawares of the southeastern part of the Reserve have been admitted to this confederacy. Not a single member is of pure Indian blood, consequently it is impossible to state the exact proportion of white and Indian in the children. An attempt was made, however, to trace back family histories two generations.

Each tribe has its own dialect. The children's ability in the use and understanding of the different dialects ranges from ability to use and understand several, to inability to use any with complete substitution of the English language. Investigation showed that forty-five per cent of the seven hundred seventeen pupils could use an Indian dialect in a conversational way. Many of the remaining fifty-five per cent could understand to a certain extent although they could not use the dialect. All of them without exception were able to speak the English language, but their ability to use the language freely was undoubtedly lower than that of white children of a corresponding social status.

These Indian pupils, measured by Chapman's Socio-economic Scale, were found to be much below his standards, the average score being thirteen against a norm of fifty-six. They attend school less regularly than the neighboring white pupils—the average attendance in percentage of total enrolment being 63.23 as contrasted with 67.14 for the

TABLE I.—MEDIAN CHRONOLOGICAL AGE IN EACH CLASS IN THE DIFFERENT SCHOOLS

School	Class[1]				
	Primer	I	II	III	IV
Ontario (rural)	7.4	8–8	10–1	11–9	13–9
Six Nations	8–5	10–6	11–1	12–4	14–1
Mohawk Institute	10–4	12–0	11–9	13–7	14–9
Mount Elgin	9–11	12–3	13–2	14–7	16–1

[1] Primer corresponds to Grade I of the American schools; Class I to Grade II; Class II to Grades III and IV; Class III to Grades V and VI; and Class IV to Grades VII and VIII.

whites. Age-grade tables show them to be retarded. Table I gives the facts for Ontario and for the Indian Schools.

From what has been said as to the nature of the subjects and their environment the selection of a series of tests which would be fair to them was a matter of some difficulty.

Practically all previous investigators had used tests involving a knowledge of English and the results obtained showed the Indians to be apparently more backward than they really were. The tests finally selected were either of the performance type or for which comparative material existed in similar studies. The tests and the number of subjects tested are given below:

I.	The National Intelligence Test, Scale A, Form I	289
II.	The Pintner Non-language Mental Test	295
III.	The Pintner-Paterson Scale of Performance Tests—the short scale of ten tests chosen by Pintner and Spaid	114
IV.	Pinter-Cunningham Primary Mental Test	59
V.	British Columbia in the Fundamentals of Arithmetic	199
VI.	Ayres Burgess Silent Reading Test	227
VII.	Ayres Spelling Scale (Long's Ontario standardization)	200
VIII.	British Columbia Spelling Test	200
IX.	Writing—Ayres Gettysburg Scale	200

The results are presented here only in the form of summary tables. Full details will be given in the report when it is published.

A. Intelligence Test Results

I. The N. I. T.

Table II.—IQ's of Various Groups of Indian Children Obtained by the Use of the N. I. T.

Group	Number tested	Q_1	Q_2, median	Q_3	QD	SD
Six Nations	141	74.0	83.7	93.8	9.9	16.2
Mohawk Institute	65	70.8	76.2	85.9	7.5	12.2
Mount Elgin	69	68.9	77.8	85.8	8.4	13.3
Total	275	71.5	79.8	89.9	9.2	15.0
Grades VII and VIII	94	75.9	85.0	94.4	9.3	13.4
Grades V and VI	103	71.8	80.5	90.6	9.4	14.5
Grades III and IV	78	64.6	74.2	82.8	9.1	15.0
Males	135	70.8	77.8	87.1	8.3	13.9
Females	140	73.0	82.4	92.6	9.8	15.7
Monoglots	153	74.1	82.4	92.4	9.2	14.8
Bilinguals	115	68.9	76.6	86.6	8.9	13.8
Six Nations (Iroquois)	183	....	81.8			
Mohawks (Bay of Quinte)	12	....	79.5			
Chippewas	28	....	77.5			
Oneidas (Thames)	28	....	73.0			
Half-bloods	12	....	85.0			
Three-quarter bloods	17	....	78.0			

As far as can be judged from Table II, the following statements seem to be justified. Indian children tested by the N. I. T. have a median IQ of about eighty. This compares with a median IQ for negroes of seventy-five and of one hundred for whites. Only ten per cent of the Indian children equal or exceed the median IQ of the white children. Residential Indian pupils are duller than day pupils and Indian girls are more intelligent by 4.6 points than Indian boys. Indian pupils who use English exclusively test 5.8 points higher than those who habitually use an Indian dialect. The more white blood the Indian pupils have the higher their intelligence as measured by this test.

II. The Pintner Non-language Mental Test

Table III.—IQ's of Various Groups of Indian Children Obtained by Use of the Pintner Non-language Mental Test

Group	Number tested	Q_1	Q_2, median	Q_3	QD	SD
Six Nations	155	86.9	101.0	115.1	14.1	18.0
Mohawk Institute	61	82.9	93.2	104.8	11.0	13.6
Mount Elgin	64	82.1	91.2	102.5	10.2	13.4
Total	280	84.3	96.9	108.7	12.2	17.6
Grades VII and VIII	93	93.0	102.0	111.8	9.7	15.6
Grades V and VI	97	85.5	94.5	107.6	11.1	17.3
Grades III and IV	84	80.4	92.0	106.2	12.9	18.8
Males	142	84.8	97.9	110.5	12.8	16.7
Females	138	83.7	96.1	106.7	11.5	18.3
Monoglots	152	86.8	100.0	110.4	11.8	17.0
Bilinguals	121	81.8	93.6	105.4	11.8	17.7
Six Nations (Iroquois)	202	....	99.3			
Mohawks (Bay of Quinte)	12	....	94.5			
Chippewas	28	....	92.0			
Oneidas (Thames)	23	....	89.0			
Half-bloods	13	....	99.0			
Three-quarter bloods	17	....	99.0			

III. The Pintner-Paterson Scale of Performance Tests (Short Scale)

Table IV.—IQ's of Various Groups of Indian Children Obtained by Use of Pinter-Paterson Scale of Performance Tests

Group	Number tested	Q_1	Q_2, median	Q_3	QD	SD
Males	51	84.6	94.4	110.3	12.9	21.1
Females	62	81.5	98.3	110.5	14.5	21.2
Total	115	82.6	96.4	110.4	13.9	21.1
Six Nations	41	93.8	107.5	119.6	12.9	21.3
Mohawk Institute	72	77.5	89.5	106.0	14.3	18.3
Monoglots	80	80.8	95.8	110.6	14.9	20.5
Bilinguals	30	87.5	100.0	109.5	11.0	22.8

IV. The Pintner-Cunningham Primary Mental Test

Table V.—IQ's of Various Groups of Indian Children Obtained by Use of the Pintner-Cunningham Primary Mental Test

Group	Number tested	Median IQ
Six Nations Schools	18	90.5
Mohawk Institute	19	80
Mount Elgin	22	66.5
Total	59	77.9
Males	24	78.3
Females	35	77.5
Monoglots	33	80.5
Bilinguals	23	68.1

If the results from the three Pinter tests are now considered, we find that they tell a different tale. Indian children on the non-language test are only three points inferior to normal white children, the monoglots, indeed, equalling the performance of other English-speaking children. On the performance test Indian children are slightly more inferior, being 3.6 points below the whites. Younger Indian children marking the pictures of the Pintner-Cunningham Primary Mental Test are given directions in English. The low median IQ of 77.9 is to be attributed in part to the handicap of language.

How do Ontario Indians compare with Indians found elsewhere and with other races? Combining Garth's ('25) findings with those for Ontario we obtain the following comparative results:

Groups	Number of cases	Median score
Twelve and thirteen years:		
Mixed bloods	15	103
Mexicans	145	85
Ontario Indians	103	80
Plains and Forest	8	76
Pueblo	46	64
Navajo and Apache	12	52
Fourteen and fifteen years:		
Mixed bloods	30	110
Ontario Indians	83	92
Mexicans	132	92
Plains and Forests	55	85
Pueblo	82	80
Navajo and Apache	19	60

On the same National Intelligence Test the one thousand fifty cases of the U. S. Indians gave a median IQ of 68.6, while the two hundred seventy-five cases of Ontario Indians gave a median IQ of 79.8.

Comparing Indians and Chinese we find that on the same Pintner Non-language Mental Test the following scores were made:

	Average IQ
513 Chinese children in Hawaii (Symonds)	99.3
280 Ontario Indian children	97.7

With the Pintner-Paterson Scale of Performance Tests (short scale) the following comparisons can be made:

	Median IQ
274 Japanese pupils in Vancouver (Sandiford)	104.5
224 Chinese pupils in Vancouver (Sandiford)	94.5
113 Ontario Indian children	96.4

Sheldon's ('24) table of the comparative intelligence of races, constructed from data in the papers of a number of investigators, remains to be mentioned. The following list shows his ranking of races on a scale of intelligence:

Nationality	Average IQ
American	100
English	100
Hebrew	98
Chinese	90
Mexican (American)	85
Indian (American)	83
Slavish	83
Italians	77
Negroes	75

The place to be assigned to the Indians of Ontario in the above array depends on which test their intelligence is to be judged. If either the Pintner Non-language Mental Test or the Pintner-Paterson Scale of Performance Tests (short scale) is the basis, they would occupy a point just below the Hebrews and above the Chinese. If they are ranked on the National Intelligence Test, their rank would be a little below the rank given to American Indians in Sheldon's table. The Chinese would also occupy a higher rank if judged by the Pintner Non-language Test. Perhaps there would be a general shifting of all the races as a result of using different tests.

B. Achievement Test Results

V. The British Columbia Test in the Fundamentals of Arithmetic

Table VI.—The British Columbia Test in the Fundamentals of Arithmetic; Median Scores in British Columbia Rural Schools, Six Nations Schools, and the Mohawk Institute

Grade	British Columbia	Six Nations	Mohawk Institute
III	22.7		
IV	38.7		
III and IV	....	28.1	
V	44.4		
VI	53.3		
V and VI	....	34.5	37.9
VII	59.4		
VIII	68.7		
VII and VIII	....	52.2	47.9

As in general intelligence, the Indians are here found to be below the level of white pupils. In their ability to perform the fundamental operations of arithmetic, Indian pupils of a given grade obtain results that are approximately equal to those obtained by white pupils who are classed one grade lower.

VI. The Ayres Burgess Silent Reading Test

Table VII.—Ayres Burgess Silent Reading: Comparison of the Median Credits in the United States, in British Columbia Rural Schools in Six Nations Day Schools, and in the Mohawk Institute

Grade	United States	British Columbia	Six Nations	Mohawk Institute
III	50	37.9		
IV	50	39.5		
III and IV	..	....	26.0	20.0
V	50	40.5		
VI	50			
V and VI	..	....	37.8	33.7
VII	50			
VIII	50			
VII and VIII	..	....	31.5	23.3

Canadian records fall below the standards in silent reading set for the United States. The Indians of Ontario are far below the rural whites of British Columbia.

VII. The Ayres Spelling Scale

Table VIII.—The Ayres Spelling Scale (List 0): Average Percentage of Correct Spellings in the United States, Ontario, Six Nations Schools, and the Mohawk Institute

Grade	United States	Ontario	Six Nations	Mohawk Institute
III	50	57		
IV	73	70		
III and IV	..	..	49	
V	84	85		
VI	92	91		
V and VI	..	..	75	71
VII	96	96		
VIII	99	98		
VII and VIII	..	..	91	83

VIII. The British Columbia Spelling Test

Table IX.—The British Columbia Spelling Test: Percentages of Correct Spellings in the United States, British Columbia, Six Nations Schools, and the Mohawk Institute

Grade	United States	British Columbia	Six Nations	Mohawk Institute
III	73	85		
IV	73	81		
III and IV	..	..	50	
V	73	81		
VI	73	86		
V and VI	..	..	47	45
VII	73	82		
VIII	73	87		
VII and VIII	..	..	61	42

Although the spelling ability of Canadian children, as a whole, tends to be superior to the standards found for the United States this superiority does not extend to Ontario Indian children. The day school Indian children spell better than the institutional children,

probably due in part to a superior mentality and partly to more extended teaching. It must be remembered that the Institute children spend half their time in farming or domestic economy of a very practical nature.

IX. Writing—Ayres Gettysburg Scale

Table X.—The Ayres Writing Scale: Norms and Medians for the United States, British Columbia, British Columbia Rural Schools, Six Nations Schools and the Mohawk Institute

Grade	United States	British Columbia	British Columbia rural	Six Nations	Mohawk Institute
III	42	48.6	48.3		
IV	46	53.3	52.0		
III and IV	..	....	....	30.2	
V	50	58.9	56.2		
VI	54	62.8	61.7		
V and VI	..	....	....	42.3	33.6
VII	58	62.9	62.0		
VIII	62	66.0	65.8		
VII and VIII	..	....	....	51.7	43.9

The comments regarding spelling apply *mutatis mutandis* for writing. Indian children make records that are two years (or grades) below those of corresponding whites.

C. Educational and Achievement Quotients

In this summary only the Educational Quotients (EQ) and the Accomplishment Quotients (AQ) in composites of scores in reading, writing, spelling and arithmetic will be given, since to print even the condensed tables for each subject would take up too much space. The trend shown by the composites is the trend shown by each subject, the Indian pupils being generally below the corresponding whites.

TABLE XI.—SUMMARY OF EDUCATIONAL QUOTIENTS IN COMPOSITES OF SCORES IN READING, WRITING, SPELLING AND ARITHMETIC OF INDIAN PUPILS IN THE SIX NATIONS DAY SCHOOLS AND THE MOHAWK INSTITUTE

Grade	Number of cases	Q_1	Q_2, median	Q_3	QD
Six Nations:					
III and IV	37	69.5	80.4	89.6	10.1
V and VI	48	72.8	83.0	93.3	10.3
VII and VIII	52	74.2	85.0	96.7	11.2
Mohawk Institute:					
V and VI	22	64.2	72.5	78.8	7.3
VII and VIII	27	73.1	78.2	83.5	5.2

TABLE XII.—SUMMARY OF ACCOMPLISHMENT QUOTIENTS IN COMPOSITES OF SCORES IN READING, WRITING, SPELLING AND ARITHMETIC OF INDIAN PUPILS IN THE SIX NATIONS DAY SCHOOLS AND THE MOHAWK INSTITUTE. THE AQ'S ARE COMPUTED ON THE BASIS OF THE NATIONAL INTELLIGENCE TEST MENTAL AGES, AND ALSO ON THE BASIS OF PINTNER NON-LANGUAGE TEST MENTAL AGES

Grade	Number of cases	Q_1	Q_2, median,	Q_3	QD
Six Nations Schools based on N. I. T. mental ages:					
III and IV.....	21	92.1	99.9	104.7	6.3
V and VI.....	30	85.6	94.0	103.1	8.7
VII and VIII	36	89.3	96.7	103.7	7.2
Based on P. N. L. T.:					
III and IV	23	68.4	78.7	95.4	13.5
V and VI	30	68.1	78.7	89.2	10.5
VII and VIII	35	74.1	81.5	87.4	6.6
Mohawk Institute based on N. I. T.:					
V and VIII	20	85.0	91.0	96.7	5.8
VII and VIII	21	91.4	94.7	99.0	3.8
Based on P. N. L. T.:					
V and VI	16	76.7	80.0	88.3	5.8
VII and VIII	21	70.8	74.1	79.7	4.5

Tables XI and XII show that the Indian children are not only doing work below the level of their chronological ages, but also below the level of their mental ages. Less than one-third of the pupils are

working up to, or above, the level of their mental capacities. Besides a possible handicap due to an actual lower level of mental capacity, the language handicap, irregularity of attendance, inferior social status and lack of heterogeneity in the environment which goes along with inferior social status, must be considered as three of the most important factors which have resulted in the inferior performances of the Indian children on the intelligence and the achievement tests. Whether or not the performances would be improved if the environmental handicaps were removed is a question for further study.

Summary

1. The language environment of the children tested is very complex. Although all of the Indian children tested had a knowledge of English, it is certain that every one of them suffered from a language handicap.

2. Not one of the subjects was of pure Indian blood, but the amount of interfusion of alien blood is indeterminable.

3. The social status of the subjects is generally lower than that of whites. School attendance is more irregular than that of white children and the amount of retardation is greater than that found in nearby rural white schools.

4. The median IQ of Indian children depends on the kind of test used. The median IQ's of all the Indian subjects according to the kind of test used are as follows:

National Intelligence Test	80
Pintner Non-language Mental Test	97
Pintner-Paterson Scale of Performance Tests	96
Pintner-Cunningham Primary Mental Test	78

The above results point to a severe language handicap on the verbal tests.

5. Indian pupils in regular day schools test higher than institutional pupils as the median IQ's show.

	N. I. T.	P. N. L. T.	P. P. S. P. T.	P. C. P. M. T.
Six Nations Schools	84	101	107.5	90
Mohawk Institute	76	93	89.5	80
Mount Elgin	78	91	.....	66

6. The females may be said to surpass the males on the whole but the differences in median IQ's is not marked.

	N. I. T.	P. N. L. T.	P. P. S. P. T.	P. C. P. M. T.
Males	78	98	94	78
Females	82	96	98	77

7. The monoglots surpass the bilinguals in all tests except the Pintner-Paterson Scale of Performance Tests.

8. The IQ seems to rise with the admixture of white blood, but the results must be interpreted with caution owing to the fact that the numbers of half-bloods and three-quarter bloods are few and also because the amount of white blood cannot be determined with accuracy.

9. The intelligence ranking of Indian subjects depends on the kind of intelligence test used. Indians rank low on the N. I. T. The use of a non-language test raises their ranking considerably. Racial median IQ's are given below:

Race	N. I. T.	P. N. L. T.	P. P. S. P. T.	P. C. P M. T.
White children	100	100	100	100
Negroes	75			
Japanese	...	...	105	
Chinese	...	99	95	
Ontario Indians	80	97	96	78
Indians in Six Nations Schools	84	101	107	90
Indians in Mohawk Institute	76	93	89	80
Indians in Mount Elgin	78	91	...	66

10. Indian pupils do not perform nearly as well as whites on the Ayres-Burgess Silent Reading Test.

11. In writing ability Indian children rank below the whites.

12. In spelling ability Indian pupils rank below the whites. They make their best showing on the Ayres Spelling Scale.

13. The Indian fourth classes (Grades VII and VIII) have a median score in arithmetic which is less than the median score of Grade VI pupils in British Columbia rural schools.

14. Both educational quotients and accomplishment quotients for Indian pupils are below normal on the average. The AQ's calculated

on mental ages from a non-language test are lower than those calculated from the N. I. T.

15. In view of the fact that both the intelligence and achievement tests used are dependent on environmental factors, any deductions from the results must be drawn carefully.

Bibliography

Arlitt, A. H.: On the Need for Caution in Establishing Race Norms. *Journal of Applied Psychology*, Vol. V, No. 2, 1921, pp. 179–183.

Bere, M.: "A Comparative Study of the Mental Capacity of Children of Foreign Parentage." Columbia University Contribution to Education, No. 154, 1924.

Brown, G. L.: Intelligence as Related to Nationality. *Journal of Educational Research*, Vol. V, No. 4, 1922, pp. 324–327.

Chapman, J. C. and Sims, V. M.: The Quantitative Measurement of Certain Aspects of Socio-economic Stauts. *Journal of Educational Psychology*, Vol. XVI, September, 1925, pp. 380–390.

Fitzgerald, J. A. and Ludeman, W. A.: The Intelligence of Indian Children. *Journal of Comparative Psychology*, Vol. VI, 1926, pp. 319–328.

Franzen, R. and Knight, F. B.: Criteria to Employ in the Choice of Tests. *Journal of Educational Psychology*, Vol. XII, October, 1921, pp. 408–412.

Freeman, F. N. : "Mental Tests." Houghton Mifflin, New York, 1926.

Garth, T. R.: Racial Differences in Mental Fatigue. *Journal of Applied Psychology*, Vol. IV, 1920, pp. 235–244.

Garth, T. R.: The Results of Some Tests on Full and Mixed Blood Indians. *Journal of Applied Psychology*, Vol. V, No. 4, 1921, pp. 359–372.

Garth, T. R.: White, Indian, and Negro Work Curves. *Journal of Applied Psychology*, Vol. V, 1921, pp. 15–25.

Garth, T. R.: A Comparison of the Mental Abilities of Mixed and Full Blood Indians on a Basis of Education. *Psychology Review*, Vol. XXIX, No. 3, 1922, pp. 221–236.

Garth, T. R.: Mental Fatigue of Mixed and Full Blood Indians. *Journal of Applied Psychology*, Vol. VI, 1922, pp. 331–341.

Garth, T. R.: The Color Preferences of Five Hundred Fifty-nine Full Blood Indians. *Journal of Experimental Psychology*, Vol. V, 1922, pp. 392–418.

Garth, T. R.: A Comparison of the Intelligence of Mexican and Mixed and Full Blood Indian Children. *Psychology Review*, Vol. XXX, 1923, pp. 388–401.

Garth, T. R.: A Review of Racial Psychology, *Psychology Bulletin*, June, 1925, pp. 343–364.

Garth, T. R., Serafini, T. J. and Dutton, D.: The Intelligence of Full Blood Indians. *Journal of Applied Psychology*, December, 1925, pp. 382–389.

Garth, T. R. and Whatley, C. A.: The Intelligence of Southern Negro Children. *School and Society*, Oct. 17, 1925, pp. 501–504.

Garth, T. R.: Race and Psychology. *The Scientific Monthly*, Vol. XXIII, No. 3, 1926, pp. 240–245.

Goodenough, F. L.: Racial Differences in the Intelligence of School Children. *Journal of Experimental Psychology*, Vol. IX, 1926, pp. 388–397.

Herrick, D. S.: A Comparison of Brahman and Panchama Children in South India with Each Other and with American Children by Means of the Goddard Form Board. *Journal of Applied Psychology*, Vol. V, No. 3, 1921, pp. 253–260.

Hunter, W. S. and Sommermier, E.: The Relation of the Degree of Indian Blood to Score on the Otis Intelligence Test. *Journal of Comparative Psychology*, Vol. II, 1921, pp. 257–275.

Jordan, A. M.: Notes on Racial Differences. *School and Society*, Vol. XVI, 1922, pp. 503–504.

Long, J. A.: The Ayres Spelling Scale in Canadian Schools. *The School*, Vol. XII, No. 4, 1923, pp. 293–304.

McCall, W. A.: "How to Measure in Education." The Macmillan Company, 1922.

Murdoch, K.: A Study of Race Differences in New York City. *School and Society*, Vol. XI, 1920, pp. 147–150.

Pintner, R. and Paterson, D. G.: "A Scale of Performance Tests." D. Appleton and Company, New York, 1917.

Pintner, R.: A Non-language Group Intelligence Test. *Journal of Applied Psychology*, Vol. III, September, 1919, pp. 199–214.

Pintner, R. and Keller, R.: Intelligence Tests of Foreign Children. *Journal of Educational Psychology*, Vol. XIII, No. 9, 1922, pp. 214–222.

Pintner, R.: "Intelligence Testing." Henry Holt, New York, 1923.

Pintner, R.: The Pintner-Cunningham Primary Test. *Journal of Educational Psychology*, Vol. XVIII, No. 1, 1927, pp. 52–58.

Pressey, S. L. and Teter, G. F.: A Comparison of Colored and White Children by Means of a Group Scale of Intelligence. *Journal of Applied Psychology*, Vol. III, 1919, pp. 277–282.

Pyle, W. H.: The Mind of the Negro Child. *School and Society*, Vol. I, 1915, pp. 357–360.

Rowe, E. C.: Five Hundred Forty-seven White and Two Hundred Sixty-eight Indian Children Tested by the Binet-Simon Tests. *Pedagogical Seminary*, Vol. XXI, 1914, pp. 454–468.

Saer, D. J.: The Effect of Bilingualism on Intelligence. *British Journal of Psychology*, Vol. XIV, July, 1923, pp. 25–38.

Sandiford, P.: "Appendix I. The Testing Programme, Survey of The School System of British Columbia, Victoria, British Columbia." 1925.

Sandiford, P. and Kerr, R.: The Intelligence of Chinese and Japanese Children. *Journal of Educational Psychology*, Vol. XVII, No. 6, 1926, pp. 361–367.

Sheldon, W. H.: The Intelligence of Mexican Children. *School and Society*, Vol. XIX, 1924, pp. 139–142.

Sunne, D.: Comparison of White and Negro Children in Verbal and Non-verbal Tests. *School and Society*, Vol. XIX, 1924, pp. 469–472.

Symonds, P. M.: The Intelligence of Chinese in Hawaii. *School and Society*, Vol. XIX, 1924, p. 442.

Walcott, G. D.: The Intelligence of Chinese Students. *School and Society*, Vol. XI, pp. 474–480.

Young, K. T.: The Intelligence of Chinese Children in San Francisco and Vicinity. *Journal of Applied Psychology*, Vol. V, No. 3, 1921, pp. 267–367.

THE EFFECT UPON RETENTION OF VARYING LENGTHS OF STUDY PERIODS AND REST INTERVALS IN DISTRIBUTED LEARNING TIME[1]

BERNICE E. CLARK[2]

The University of Chicago

Interest in distributing time for learning has waned with the failure to find a schedule optimal for all material, leaving the student to apply the general fact of its advantage as best he can.

If learning distributed at intervals over a period yields a fuller retention than when the time required for it is concentrated at one sitting, some factor or group of factors of the learning schedule must determine this advantage. On analysis but three such factors appear; the length of the whole schedule, the lengths of the learning periods and the lengths of the rest intervals, all of which condition the number of study periods. Now should one of these variables be found to be the factor which most influences the retention, or failing this, could the relative amounts contributed by each be determined, general principles might evolve, by means of which learning schedules suitable to the kind and length of material could be constructed.

This investigation studied the effect upon a delayed retention made by varying the lengths of the study periods and the lengths of the rest intervals between. Specifically it sought to discover whether the learning periods of equal or decreasing lengths, or rest intervals of equal or increasing lengths produced a fuller retention three weeks after the learning.

Early investigators seeking an optimal schedule distributed learning time in equal portions with equal rest intervals between. Not until 1914 did Lyon[3] advocate doubling the length of successive rest intervals as an aid to retention. Learning periods with lengths successively decreasing had been suggested by Cummins' investigation[4] of 1916

[1] A study conducted under the direction of Professor Frank N. Freeman, The University of Chicago.

[2] Graduate student.

[3] Lyon: The Relation of Learning of Material to the Time Taken for Learning and the Optimal Distribution of Time. *Journal of Educational Psychology*, Vol. V, 1914, p. 160.

[4] Cummins, Robert A.: "Improvement and the Distribution of Practice." New York: 1919, p. 9. Teachers College, Columbia University Contributions to Education, No. 97.

and the writer's earlier experience in using the Ebbinghaus Saving Method for learning sonnets. Cummins found that practice distributed in learning periods of the decreasing type, separated by rest intervals which grew gradually longer, showed greater improvement than when distributed regularly over a schedule of the same length. He, however, was uncertain whether the study periods or the rest intervals produced this difference. Moreover, his irregular time allotments were so varied (even reversed in order) that one questions whether the same factors uniformly determined the results. Further study of irregular time allotments for learning schedules was deemed promising.

The present experiment was carried on in a senior high school during the school year of 1925–26. The plan arranged for four schedules, A, B, C, and D, constructed to test the respective contributions of learning and rest periods. A and B had like learning periods, all of equal lengths; similarly, C and D had learning periods successively decreasing in length. The rest intervals of A and C were alike and equal, while those of B and D increased in duration.

Many of the details of the plan grew out of the experience in the preliminary study, the method of which paralleled this.

Material.—In that investigation English sonnets served as the material to be learned. The choice was first made on the assumption that two sonnets are more comparable in form, unity, and content than six stanzas of narration and six stanzas of description, selected from a single poem like "Don Juan." However when the time needed by the sixteen subjects to learn two Miltonian sonnets was compared, the correlation was $78 \pm .07$. This is less of a difference than Ebbinghaus found in his own learning[1] of seven groups of six stanzas from "Don Juan." Sonnets of Keats, Milton, Shakespeare and Wordsworth were used; eight in all, four demanded by the plan with an alternate for each.

Time Allotments for the Schedules.—Time allotments commensurate with the length and difficulty were then fixed. Adults in the first study had on the average needed about twenty-four minutes for learning a sonnet, hence thirty minutes were deemed sufficient for adolescents.

The number of learning periods together with the duration of the decreasing learning periods were determined by the experience of

[1] Ebbinghaus, Herman: "Memory: A Contribution to Experimental Psychology," Translated by H. A. Ruger and Clara E. Bussenius. New York: Teachers College, Columbia University, 1913.

Ebbinghaus and the sixteen subjects of the first experiment in learning sense material by the Saving Method. In both experiments, when learning was brought to the point of an errorless recall at sittings separated by intervals varying from one to seventy-two hours, two results obtained. In the first four or five such study periods were sufficient to learn the material so that it could (twenty-four hours later) be repeated without recourse to copy. Based on this agreement four learning periods were assigned to each schedule. A second finding common to both showed the duration of the study periods decreased in a geometric progression. In both investigations the ratio wavered between one-third and one-half. A convenient interpretation of this divided the thirty minutes of schedules C and D into successive learning periods of fifteen, eight, five and two minutes.

No interval between learning periods was less than twenty-four hours, since it was difficult for a student to arrange for more than one appointment per day. In order to double the length of successive rest intervals of the irregular schedules, as well as to avoid appointments on non-school days, each schedule covered eight days.

Time allotments for the learning periods and rest intervals of each schedule follow. In Schedules A and B, each of the four learning periods was 7½ minutes long, while in C and D they were fifteen, eight, five and two minutes respectively. The first two of the three rest intervals of Schedules A and C were each forty-eight hours in length; the last, seventy-two hours. The rest intervals of B and D were twenty-four hours, forty-eight hours and ninety-six hours, respectively. The additional twenty-four hours in the last rest interval of all schedules was an allowance made for Sunday.

Subjects.—Thirty-two senior high school students completed the four schedules; eighteen of these (senior girls) were members of the writer's advisory group and volunteered their services; the remaining fourteen boys and girls were chosen from the writer's classes. The subjects were placed at random in four groups of eight each. The following arrangement (suggested by Dr. Gates[1]) of groups, sonnets, and order of learning sought to control the factors of practice, unequal difficulty among the sonnets, and individual differences among the subjects.

---

[1] Gates, Arthur I.: Recitation as a Factor in Memorizing. *Archives of Psychology*, Columbia University, No. 40. Edited by R. S. Woodworth. New York: The Science Press, 1917.

TABLE I.—ARRANGEMENT OF GROUPS, SONNETS AND ORDER OF LEARNING IN THE FOUR SCHEDULES

Schedules	Order of learning			
	First	Second	Third	Fourth
A	Group I Sonnet I	Group II Sonnet IV	Group III Sonnet II	Group IV Sonnet III
B	Group III Sonnet III	Group IV Sonnet II	Group I Sonnet IV	Group II Sonnet I
C	Group II Sonnet II	Group I Sonnet III	Group IV Sonnet I	Group III Sonnet IV
D	Group IV Sonnet IV	Group III Sonnet I	Group II Sonnet III	Group I Sonnet II

Procedure.—Since each subject worked alone under the observation of the writer, the procedure was simple. Before he began, a student was warned of the exacting demands of the experiment and if he felt the demands too taxing he was not urged to undertake the work. Upon his agreement to take part in the experiment, a time and place were then agreed on, when and where the subject could best meet all appointments for the experiment.

On his first appearance a typewritten copy of the material was given him. From a duplicate the experimentor read the sonnet aloud, offering at the end to explain any difficult allusion. No effort was made to control the learning; the experimentor did, however, recommend the whole-part method and attempts at repetition as aids. The observer made notes on the behavior of the subject, who in turn reported at the end of each sitting, his progress, his method of learning, and what other impressions he desired.

At the close of the fourth study period the student wrote what he had learned. From five to seven minutes (based on the average time needed in the preliminary study) were granted for this record, hereafter referred to as the learning test. Twenty-one days after the learning test the subject was given ten minutes in which to write what he could recall of the sonnet previously learned. This was called the retention test. His retention score was computed by subtracting the number of errors in the learning test from those in the retention test. When the

retention test of one sonnet was completed, a subject began, as soon as possible, his subsequent schedule.

Scoring.—The problem of scoring the tests had been met in the earlier study. In both investigations the writer was less concerned with the actual amount of each sonnet remembered than with the relative amounts of the four sonnets retained. Any arbitrary scale of errors which considered degrees of retentiveness would serve. To rate the penalties for the four most prevalent errors, the following key was devised:

	Error
1. For each word omitted	1
2. For each wrong word	1
3. For each word of a misplaced phrase, regardless of the extent of its misplacement	½
4. For each synonym which retained the original rhythm	½

In cases where the key did not suffice, the scorer's judgment must meet the exigency. Since all scoring was done by the writer, variability in that direction, however, was reduced.

Quantitative Results.—In writing his sonnet, no subject required more than five minutes for the learning test and seven minutes for the retention test, time limits allowed by the plan. With the speed element eliminated the scores indicate only the amount of any sonnet forgot.

An examination of the individual scores in Table II admits none of the four schedules consistently the best. The majority of subjects, however, retained sonnets learned on Schedules C and D better than those learned on A and B. If the total errors for each of the four quarters be found, the superiority of Schedules C and D becomes more apparent. In the total errors made by the thirty-two subjects on each schedule this tendency is pronounced. A's and B's total errors are identical, an implication that the rest intervals exercise little influence on the retention. This inference is not however borne out by the difference in the totals of C and D.[1] A's total errors exceed

[1] To further test the contribution of rest intervals to retention, a supplementary experiment arranged for greater differences among the lengths of the increasing rest intervals. Twenty-four subjects learned sonnets on two schedules, each covering eight days. Both had learning periods of decreasing duration (like C and D); in one the rest intervals were equal, in the other they were eight hours, twenty-four hours and ninety-six hours. Arrangements and procedure followed the present study. In the retention test, twenty-eight days after the learning, the difference in the total errors of the two schedules was but two. This agreement with the identical total errors of Schedules A and B (Table II), and the data obtained from

TABLE II.—ERRORS IN THE RETENTION OF THIRTY-TWO SUBJECTS RANKED IN ORDER OF THE TOTAL NUMBER OF ERRORS MADE ON THE FOUR SCHEDULES BY EACH SUBJECT

Subject	Errors made on schedules				
	A	B	C	D	Total
I	0	0	0	0	0
II	0	0	0	.5	.5
III	0	0	1	−.5	.5
IV	0	0	0	1	1
V	.5	.5	0	1	2
VI	5	.5	.5	0	6
VII	0	−7	9	4	6
VIII	1	3.5	1.5	.5	6.5
IX	6	1.5	.5	.5	8.5
X	−1	12.5	.5	3	15
XI	3	0	10	3	16
XII	4.5	0	11	1.5	17
XIII	−4.5	9.5	3	12.5	20.5
XIV	13.5	1	8.5	0	23
XV	17	3.5	.5	4	25
XVI	16	19.5	0	.5	36
XVII	1.5	−2	35.5	1.5	36.5
XVIII	20	23	2	.5	45.5
XIX	1.5	18.5	32	3	55
XX	24.5	11.5	35	0	71
XXI	60	9	.5	8	77.5
XXII	1.5	87	17	4.5	110
XXIII	0	71.5	−7	47	111.5
XXIV	71.5	8	21.5	13.5	114.5
XXV	1	15	25.5	74.5	116
XXVI	67	26.5	7.5	19	120
XXVII	69.5	27	31.5	3.5	131.5
XXVIII	55.5	37	60	15	167.5
XXIX	50	50	75	.5	175.5
XXX	25	80	36.5	60	201.5
XXXI	65	64.5	46.5	34	210
XXXII	51	54	31	83	219
Total	625.5	625.5	496	399	2146

the preliminary study implies the rest intervals affect retention but little. Could it be that the advantage ascribed by Lyon to the doubling of successive rest intervals was due to the increased length of the entire schedule?

D's by 226.5, a difference of fifty-six per cent, and also exceeds C's by one hundred twenty-nine errors, more than one-half the first excess.

In the main the results infer that the superiority of Schedules C and D is due to the decreasing lengths of the study periods and particularly due to learning periods with lengths decreasing in a geometric progression.

From the notes made by each subject, chance factors such as fixed habits of learning or an emotional predilection for certain subject-matter were found to explain the more striking deviations (subjects XVII and XXIII) from the general trend. Even with the principal variables under control, an unwavering advantage of a single schedule is perhaps too much to expect. On the whole the data here are quite as consistent as those of Cummins.

Combination of Schedules $(A + B)$ *and* $(C + D)$.—The similarity in totals of Schedules A and B and in the totals of C and D suggests a close relationship of schedules having like learning periods. If an individual's errors on Schedules A and B are added and a similar combination of the scores of C and D made, such pairing ignores the influence of the rest intervals and tends, besides, to reduce the hazard that a chance factor might upset a single schedule. In Table III the data thus combined are ranked in order of the sum of errors made by an individual on the paired Schedule $(C + D)$. In this arrangement, only two subjects do better on the pair $(A + B)$. If summaries of the errors in each quarter be made, individual variation is entirely submerged. And in the total errors the pair $(A + B)$ exceeds the pair $(C + D)$ by three hundred fifty-six errors, an excess of 39.8 per cent. It will readily be found that no other pairing of schedules yields so large a difference or so evident a consistency.

The outcome of the pairing not only supports Cummins' findings but offers an answer to his query: To which factor, the decreasing learning periods or the increasing rest intervals, is the advantage of the Reducing Schedule due. The answer is summed up in the following conclusions:

1. The duration of the learning periods has more influence on retention than the duration of the rest intervals.

2. Study periods with lengths decreasing in a geometric progression, when compared with study periods of equal length, reduce considerably the rate of forgetting.

These findings are, however, but a block in the pattern. Learning periods with lengths decreasing in a geometric progression should be adapted to material of varying lengths and difficulty, to determine

TABLE III.—ERRORS IN THE RETENTION OF THIRTY-TWO SUBJECTS RANKED ACCORDING TO THE SUM OF THEIR ERRORS MADE ON SCHEDULE (C – D)

Subject	Errors made on schedules	
	(C – D)	(A – B)
I	0	0
II	.5	0
III	.5	0
VI	.5	5.5
XVI	.5	35.5
IV	1	0
V	1	1
IX	1	7.5
VIII	2	4.5
XVIII	2.5	43
X	3.5	11.5
XV	4.5	20.5
XIV	8.5	14.5
XXI	8.5	69
XII	12.5	4.5
VII	13	−7
XI	13	3
XIII	15.5	5
XXII	21.5	88.5
XXVI	26.5	93.5
XIX	35	20
XX	35	36
XXIV	35	79.5
XXVII	35	96.5
XVII	37	−.5
XXIII	40	71.5
XXVIII	75	92.5
XXIX	75.5	100
XXXI	80.5	129.5
XXX	96.5	105
XXV	100	16
XXXII	114	105
Total	895	1251

for what limits, if any, this principle does not hold. In a similar way the respective contributions to delayed retention made by the duration of learning periods and the length of the entire schedule should be studied.

MATCHING ABILITY AS A FACTOR IN FIRST GRADE READING

NILA BANTON SMITH

Supervisor of Research, Detroit Public Schools

Reading is a perceptual process, consisting of the recognition and interpretation of symbols. One of the first steps in any type of perceptual learning is that of discriminating between sensations. Applied to beginning reading then, one of the basic elements in gaining a percept of a new word is that of obtaining a clear cut visual impression of this particular word as being distinctive from other words. We must learn to discriminate between this word and all other words in order that we may recognize it as being the same word every time that we see it in the future. General ability in word recognition cannot progress very far until pupils have developed the particular ability of making visual discriminations at least to the extent of being able to match word forms.

Some children possess this ability when they come into the first grade. Many do not. Even a casual observer in a first grade room will have noted instances in which the children have difficulty during the early weeks when directed by the teacher to find the same word twice when it appears in two different places upon the board or upon cards in the chalk ledge. Since matching performance plays such an important part in learning to read, a question arises as to the relationship between matching ability and reading ability. Does the child who can match well at the beginning of the term attain a greater success in reading than the one who has difficulty in matching?

In order to obtain data bearing on this problem, the series of investigations described in this article were conducted with beginning first grade children in Detroit schools during the school year of 1926–27. The first step taken toward procuring such data was that of preparing material for the investigation. Each of 12 teachers printed two complete sets of lower case letters placing each upon a separate card. A rubber printing outfit was used so that all letters would be uniform in size and shape, and so that all like letters would look *exactly* alike. One of the completed sets of cards was then placed in a certain disarranged order on a low table, the other was arranged in true alphabetical order in pack form and placed on the teacher's desk.

Having prepared and arranged the material, the teacher in each case proceeded to conduct the study by calling to her one pupil at a

time, giving to him the "a" card out of her pack and asking him to find one exactly like it on the table among the disarranged cards. At the end of two minutes the child was given a score of right or wrong according to the success of his attempt, and was asked to match the next letter in the teacher's pack and so on, till he had had a trial at matching all of the letters in the alphabet. Each teacher chose ten beginning Grade I pupils for use in this particular phase of the investigation, and in each case, she selected three X's, four Y's and three Z's,* the total number who matched lower case letters being 120. The experiment took place during the first week of the term before the children had had any matching experience in connection with their reading activities.

At the end of 12 weeks the reading ability of these same pupils was tested by administering to them the "Detroit Word Recognition Test." The results of this reading test were then compared with the scores obtained by giving the matching test during the first week of the term, and the coefficient of correlation computed. Table I shows the data revealed by this computation.

TABLE I.—CORRELATION BETWEEN MATCHING ABILITY AND READING ABILITY
Matching Scores, Lower Case Letters, First Week, Reading Scores, Recognition Test, Twelfth Week

GRADE	CASES	COEFFICIENT OF CORRELATION
B1	200	.87

It will be seen in the table above that the resulting coefficient of correlation was .87 which, of course, is very high. This high correlation between matching ability and reading ability is significant from two standpoints.

First, it has significance from the standpoint of its prognostic value as a determinant of reading maturity. Is it not possible that a test based on word and letter matching might furnish us with an indication of reading readiness? In other words, would such a test administered at the beginning of the term enable us to predict which children are ready for reading and consequently will be able to take it up successfully and which ones are too immature at that time to enter into the formal reading activities? The first grades throughout the country are clogged each year with pupils who fail simply because they are not

* In the Detroit school system pupils are classified into X, Y, and Z groups according to their intelligence. The X children are those having superior intelligence, the Y's are those having average intelligence, and the Z's are those having an intelligence below the average.

mature enough mentally to engage in the formal reading activities as organized by our present classroom methods. Any instrument which would enable us to predict reading readiness would be tremendously useful. The results of the investigation described above would lead us to believe that some measure of matching ability would be highly prognostic of possible success in reading.

The second possibility which presents itself in this connection is whether or not a greater emphasis upon matching activities in the kindergarten might serve better to prepare children to take up the study of Grade I reading. Whether or not ability to match things in general would transfer to ability in matching symbols is, of course, another line of experimentation. The results of this experiment point toward an affirmative answer to this possibility, but we must await further investigation in order to obtain conclusive evidence on this point.

Several other interesting details resulted from a series of matching studies which were conducted during this general investigation, some of which will be presented.

In tabulating the scores resulting from having the children match lower case letters as described above, it seemed desirable to arrange the letters in the order of their relative difficulty. This data is shown in Table II.

TABLE II.—RELATIVE DIFFICULTY IN MATCHING LOWER CASE LETTERS

	PER CENT OF ERRORS		PER CENT OF ERRORS
b	.341	x	.066
p	.281	y	.066
q	.225	k	.058
d	.216	t	.058
r	.141	z	.050
h	.133	l	.050
f	.133	a	.041
i	.125	c	.041
j	.125	w	.041
n	.125	o	.041
u	.108	m	.033
e	.066	s	.033
v	.066	g	.025

In examining this table we find b, p, q, and d, to be by far the most difficult letters for children to match. This may be accounted for in that each one is composed of a loop and a stem, and children do not seem to notice on which side the loop occurs, or whether the stem extends upward or downward. r, h, f, i, j, n, and m, are next in

difficulty, each being missed by around 13 per cent of the pupils. e, v, x, y, k, t, z, and l, occur next in frequency ranging from 6½ per cent to 5 per cent. a, c, w, o, m, s, and g, were missed only by 4 per cent or less of the children, so, of course, are the ones which offer the least amount of matching difficulty.

Ability to match lower case letters assumes special importance in the Detroit method of reading as it is through the application of this ability that pupils are able to use a dictionary index in locating in a pictured dictionary new words which they do not recognize in their reading lessons. Consequently, teachers find it necessary to give considerable practice in matching lower case letters during the early part of the term. In the past this practice has not been given in any systematic manner. A child has been called upon as frequently to match one letter as another, while as a matter of fact, some letters are very difficult to match and others are very easy to match. Provided with the data above, the teacher has information as to the relative difficulty of the various letters, and consequently can distribute her practice accordingly.

In connection with this problem it seemed advisable to find the likenesses and differences of the X, Y, and Z groups, so that the teachers may deal with each level of intelligence according to its special strengths and weaknesses. Consequently, the data for each group was tabulated separately and arranged as shown in Table III.

TABLE III.—X, Y, Z DIFFERENCES IN MATCHING LOWER CASE LETTERS

February, 1927, B1 Grade, First Week of Term, 125 Cases

X's Missed in Per Cent		Y's Missed in Per Cent				Z's Missed in Per Cent			
b	22	b	38	y	11	b	50	l	13
p	20	p	36	v	9	d	43	t	13
d	15	d	22	h	7	q	37	o	10
q	15	q	22	t	7	p	30	m	10
h	5	i	22	k	4	f	27	w	10
f	5	f	18	a	2	i	27	o	10
n	5	j	18	c	2	r	27	x	10
c	2	r	18	l	2	j	23	z	10
m	2	u	16	m	2	h	20	e	7
s	2	n	11	s	2	n	20	g	7
t	2	x	11	z	2	k	17	s	7
u	2					u	17	v	7
						a	13	y	7
Not Missed a, e, g, i, j, k, l, o, r, v, w, x, y, and z		Not Missed e, g, o, w							

In this table we see the letters arranged in the order of their matching difficulty for X's, Y's and Z's respectively. We find b, p, d, and q occurring with the highest frequency in each group. In the remainder of the discussion I shall refer to these letters as the "matching demons." In the case of the X's we find several letters which were not missed by any children at all. a, e, g, i, j, k, l, o, r, v, w, x, y, and z were correctly matched by every X pupil who was given a trial. In the Y group we find only four letters which were not missed by any child, e, g, o, and w. These four are included within the group which was correctly matched by all X's. In the case of Z's we find that every letter was missed by some of the children, 7 per cent being the smallest per cent of error for any one letter.

These data suggested the possibility of marking off certain areas of control for each of the X, Y, and Z groups, that is, placing letters which are very difficult for any one group to match in the area of "Low Control," signifying that these letters are the ones to which we must devote the most attention; then allotting the middle letters in the table to the area of "Medium Control," indicating that they are the ones to which a medium amount of attention must be given; and finally allotting those which are missed by a small percentage or none of the pupils in the area of "High Control," as an indication that they do not require any special practice. In marking off the control areas, I decided to place all letters missed by 15 per cent or more of the children in the area of "Low Control," those missed by between 5 and 15 per cent in the area of "Medium Control," and those missed by less than 5 per cent in the area of "High Control." Working upon this basis, Table IV was prepared.

TABLE IV.—CONTROL AREAS IN MATCHING LOWER CASE LETTERS
February 1927, B1 Grade, First Week of Term, 125 Cases

X	Low Control	Medium Control	High Control
Y	Low Control	Medium Control	High Control
Z	Low Control	Medium Control	

b p d q i f j r u n x y v h t k a e l m s z c w o g

In the light of this interpretation we find that the area of low control for the X's covers the letters, b, p, d, q; while their area of medium control covers i, f, j, r, u, n, x, y, v, and h; and their area of high control covers t, k, a, e, l, m, s, z, c, w, o, and g. In case of the Y's we find the area of low control is about twice as long as that for the Z's covering b, p, d, q, i, f, j, r. Their area of medium control is not as long as that of the X's, covering only u, n, x, y, v, h, and t. Their area of high control is very nearly as long as that of the X's covering k, a, e, l, m, s, z, c, w, o, and g. In the case of the Z's we find an area of low control which is nearly twice as long as the Y's or four times as long as the X's, covering b, p, d, q, i, f, j, r, u, n, x, y, v, h, t, k, and a. Their area of medium control covers c, l, m, s, z, e, w, o, and g. They have no area of high control.

This allotment of control areas is but my own interpretation of the situation and possibly the divisions I have made are not the wisest ones to make, but this treatment at least serves to throw into relief the general degree of difficulty of different groups of letters of each of the grades of intelligence. Provided with such information, a teacher has a guide which should enable her to deal more intelligently in teaching letter matching to each of her groups.

In this connection it also seemed desirable to find out if there was any relationship between ability to match lower case letters and intelligence. As a rough means of determining this relationship, I computed the median of the matching scores of each group separately. The results are shown in Table V.

TABLE V.—MEDIAN MATCHING SCORES OF THE DIFFERENT INTELLIGENCE GROUPING

February 1927, Bl Grade

INTELLIGENCE CLASSIFICATION	MEDIAN MATCHING SCORE LOWER CASE LETTERS
X Group (30 Cases)	25.8
Y Group (40 Cases)	24.2
Z Group (30 Cases)	20.7

In referring to this table we find that the X's made a median matching score of 25.8, while the Y's earned a median of 24.2 which is 1.2 points less than the X's. The Z's made a matching score of only 20.7, which is 3.5 point less than the Y's. The wider difference between the Y's and Z's may be accounted for in part by the fact that motor skill entered into the matching activity to a certain extent as the child took the card in his hand in each case and passed to the

table where he proceeded to carry out the mental activity of matching. Possibly it took the Z longer to take the card in his hand and pass to the table than it would take an X or Y and in this way he may have been slightly hampered in his total matching time.

The general medians as they stand, however, do show a correspondence between matching ability and different levels of intelligence as indicated by the X, Y, and Z grouping.

So far the discussion has been concerned only with ability to match lower case letters. This phase of the study is especially significant to teachers who must teach children to use an alphabetical index. It might also be of value to teachers of spelling and writing, but as far as first grade reading in general is concerned, it is not of much importance, because we no longer teach reading by the alphabet method but rather by word or phrase methods. Consequently a point of greater significance as far as general reading is concerned is that of determining the effect which the different letters have upon matching when combined into groups as they would appear in words. In an effort to get some light on this problem I prepared nine non-sense syllables involving some of the letters which were difficult to match, and some which were easy to match and combining these letters in different arrangements as you see in Table VI.

TABLE VI.—PER CENT OF ERRORS IN MATCHING LOWER CASE COMBINATIONS
February 1927, B1 Grade, First Week of Term, 100 Cases

		Per cent of errors for each syllable	Average per cent of errors for each group
Hard letters at beginning and end..	bid	21.3	19.2
	dip	19.6	
	piq	16.8	
Easy letters at beginning, hard letters at end.....	yid	18.5	17.5
	gip	17.4	
	ziq	16.8	
Hard letters at beginning, easy letters at end.....	diy	16.4	15.3
	pig	16.4	
	qiz	13.1	

In referring to this table we find that the middle letter "i" is kept constant all the way through, but in the first three combinations we have a letter which is very difficult to match at both the beginning and end of the syllables; in the second group we have an easy letter at the beginning of each combination, and a hard one at the end; while in the third group the reverse is true, each syllable beginning with a hard letter and ending with an easy one.

The same procedure was used in studying the matching of these syllables as that used in the study with lower case letters. Had the results come out as we at first thought we might expect we would find that those syllables both beginning and ending with hard letters would be very much more difficult for children to match than those which had one easy letter either at the beginning or end, but as a matter of fact the figures do not show that this was always the case. Take the syllable "piq" for instance, which both begins and ends with a hard letter; this combination was missed by 16.8 of the children, while "yid" which involves one easy letter and one hard letter was missed by 18.5 per cent of the children. So we see that a word both beginning and ending with a matching demon is not necessarily harder to match than one which begins with an easy letter and contains only one demon at the end.

Now let us see if it makes any difference whether the demon comes at the beginning of a word or at the end. In the three cases which begin with easy letters and end with a demon, we have percentages of error as follows: 18.5, 17.4, and 16.8 as compared with 16.4, 16.4 and 13.1 in the cases in which the syllables begin with hard letters and end with easy ones. This shows a slight decrease in errors in favor of the latter arrangement. This evidence points toward the conclusion that it is the last part of a word rather than the first part which gives it its distinction in form from the child's standpoint of observation, but the number of cases is too small and the differences are too slight to warrant any definite statement to this effect.

In considering the average per cent of errors, we find that syllables both beginning and ending with demons have a slightly higher percentage of error than those in which only one demon appears. We also find the lowest average of errors occurring in cases in which the demon is at the beginning and the easy letter at the end. The difference between the first and second averages, however, is only 2 per cent and the difference between the second and third averages is only 1½ per

cent. These differences are so slight as to have little significance, the general conclusion being that difficulties in matching single letters do not carry over to any appreciable extent in matching words. This conclusion supports the evidence of other investigators to the effect that children do not perceive units by letter but rather are they influenced by certain characteristics of word wholes.

So much for lower case letters and combinations. The second general topic of investigation was concerned with problems in matching capitals. All people associated with little children are well aware of the difficulty which they experience in recognizing an old word when it first appears in its capitalized form. The capital at the beginning changes the word form to the extent that it often appears to the child as an entirely new and different word. Again, it seems that a teacher might profit by having placed in her hands some specific information regarding difficulties in matching capitals.

In investigating this problem, the same general procedure was used as that described in the study of lower case letters. The results of attempting to find the relative difficulty of matching capitals with capitals is shown in Table VII.

TABLE VII.—RELATIVE DIFFICULTY IN MATCHING CAPITALS WITH CAPITALS
February 1927, B1 Grade, First Week of Term 50 Cases

	PER CENT OF ERRORS		PER CENT OF ERRORS		PER CENT OF ERRORS
Q	25	G	13	A	0
B	23	K	13	C	0
D	22	N	13	F	0
J	22	M	13	H	0
Y	22	P	13	I	0
Z	22	T	13	O	0
R	22	V	13	S	0
U	22	W	13	X	0
E	13	L	4		

Q was found to have caused the highest percentage of difficulty. B, D, J, Y, Z, R, and U occur next in difficulty; E, G, K, N, M, P, T, V, and W constitute the next group; L is next, and A, C, F, H, I, O, S, and X were not missed by any children.

It is interesting to note that, in general, it is easier for children to match capitals than lower case letters. The range of the per cent of errors for the lower case letters extending from 34.5 per cent to 2.5 per cent, whereas the range of errors for the capitals extends from 24 per cent to 0 per cent for the last six letters which offered no difficulty what-

ever. The greater ease with which children match capitals may be attributed in part to the fact that capitals are larger and more distinctive in form than lower case letters. It is also true that blocks and educational toys with which children are provided before they come to school are in most cases marked with the capital letters, so perhaps they are more accustomed to seeing the capitalized forms.

While it is evident that capitals are easier to match with one another than lower case letters are, when children are asked to match lower case letters with their corresponding capitals, we find quite a different story, as revealed in Table VIII.

TABLE VIII.—RELATIVE DIFFICULTY IN MATCHING LOWER CASE LETTERS WITH CAPITALS

February 1927, Bl Grade, First Week of Term, 50 Cases

	PER CENT OF ERRORS		PER CENT OF ERRORS		PER CENT OF ERRORS
q-Q	40	i-I	28	w-W	10
r-R	38	n-N	28	c-C	10
g-G	36	t-T	26	z-Z	6
b-B	34	d-D	24	s-S	4
f-F	34	m-M	24	v-V	4
a-A	32	j-J	22	k-K	2
e-E	32	y-Y	18	o-O	2
h-H	32	u-U	12	x-X	2
l-L	30	p-P	10		

In this table which shows the relative difficulty of matching lower case letters with capitals, we find that f, r, g, b, and q were missed by from 40 to 34 per cent of the children; a, e, h, l, i, n, t, d, m, j, and y are missed by from 32 to 18 per cent of the children; u, p, w, c, and z are much easier because of the similarity of the two forms, and x, v, k, o, and x present only a very slight difficulty, which can be readily understood because in these cases the two forms are practically identical.

As to the general difficulty of matching lower case letters with capitals we find the percentage of error ranging from 40 to 2. It is evident, then, that it is more difficult for children to match lower case letters with their corresponding capitals, than it is for them, either to match lower case letters with lower case letters, or capitals with capitals.

Since children rarely engage in the isolated activity of matching lower case letters with their corresponding capitals, this data in itself is not significant. The real situation with which pupils are continually

confronted is that of meeting a word for the first time in its capitalized form, which previously they have only met in its lower case form. So the point with which we are most concerned is whether these letter difficulties carry over in recognizing the capitalized form of word wholes. Table IX shows the results of a study on this phase of the problem.

TABLE IX.—ERRORS IN RECOGNITION OF WORDS BEGINNING WITH CAPITALS AS COMPARED WITH THOSE BEGINNING WITH SMALL LETTERS
50 Cases, B1 Grade, End of First Month

Easy				Medium				Hard			
Lower case words	Missed	Upper case words	Missed	Lower case words	Missed	Upper case words	Missed	Lower case words	Missed	Upper case words	Missed
out	0	Out	0	tree	0	Tree	18	get	1	Get	34
wolf	0	Wolf	1	milk	0	Milk	14	red	0	Red	30
cat	0	Cat	0	jam	0	Jam	12	fox	0	Fox	32
Total.....	0	.....	1	....	0	.....	44	....	1	....	96

In order to obtain this data, I selected three words which began with the capitals which we found were the easiest for children to identify with their corresponding lower case forms. Then I selected three words which began with letters of medium difficulty, and then three more which began with the most difficult capitals. Fifty children were taught to recognize the lower case forms of all these words. At no time during the teaching process did they have an opportunity to see the capitalized form. After the teacher thought the pupils knew the words one hundred per cent, each child was taken into another room by himself, three lower case words were placed on the chalk ledge in each case, and the child was asked to read them. After reading the lower case forms, these cards were collected and the capitalized forms of the words placed in the chalk ledge in their stead. The child was then asked if he knew any of these words. The results are recorded above. In the case of *out*, *wolf*, and *cat* which words began with letters easy to identify with their capitals we find that all of the words are recognized in their lower case form, and all of them were recognized in the capitalized form excepting *wolf*, which was missed by one child. In noting those beginning with letters of medium difficulty, we find all of the children recognized *tree*, *milk*, and *jam* in

their lower case form, but that 18 missed *Tree* in its capitalized form, 14 missed *Milk* and 12 *Jam*. In examining the results of those beginning with the hard letters, we find only one error in recognizing the lower case forms, whereas 34 children missed *Get* in its capitalized form, 30 missed *Red*, and 32 missed *Fox*, showing that the capital in each case made the form appear as an entirely different word to about 60 per cent of the children. The results show only one error in the case of the easy letters, 44 in the case of the medium letters, and 96 in the case of the hard ones. From this data it is evident that the letter difficulties presented in Table VII do carry over in recognizing the capitalized forms of word wholes.

The experiment as a whole is not a pretentious one. Its greatest merit probably lies in the fact that its results are of such a concrete practical nature that they may be placed directly in the hands of teachers for immediate application in meeting their matching problems.

RECIPROCAL RATINGS: A COMPARISON OF THE JUDGMENTS OF A TEACHER AND PUPILS IN ESTIMATING ABILITY IN UNITED STATES GEOGRAPHY

RICHARD S. UHRBROCK

Cornell University

I. Introduction

In standardizing an educational test, the opinion of the teacher usually is rated as "expert" by the test constructor. In most cases no attempt is made to determine whether or not the teacher has what Woodworth calls "good judgment"—"the taking of all relevant stimuli into account, and letting the response be aroused by the combination."[1] It has been demonstrated that teachers' estimates vary considerably in ranking the same school material at different times.[2] Such evidence suggests the possibility that grades and ranks assigned by the average teacher may be unreliable criteria in standardizing educational tests.

We tend toward reliable estimates of proficiency in school subjects when we secure several independent judgments from the teacher and average those judgments. Rugg has further emphasized this point by stating that we should "*Obtain a mass judgment from good judges.*"[3] Immediately we are confronted with the problem of discovering "good" judges. It has been suggested that school children may be able to rate their classmates in a given subject with greater accuracy than their teacher is able to show. The pupils are supposed to be able to distinguish between the ones who prepare their assignments thoroughly and the ones who rely upon last minute assistance from others. Shelton[4] experimented with mutual ratings "where each individual in the entire personnel of a large office rated every other individual by means of a secret ballot." His report suggested the problem which has been answered tentatively in this paper.

---

[1] Woodworth, Robert S.: "Psychology, A Study of Mental Life." Henry Holt and Co., New York, 1921.

[2] Starch, Daniel: "Educational Psychology." The Macmillan Co., New York, 1927.

[3] Rugg, Harold O.: Is the Rating of Human Character Practicable? *The Journal of Educational Psychology*, Vol. XIII, February, 1922.

[4] Shelton, H. W.: Mutual Rating. *Bulletin of the Taylor Society*, Vol. V, No. 2, April, 1920.

II. Statement of the Problem

This study is an attempt to answer the following question: *Can pupils estimate the relative standing of their classmates in a school subject with greater accuracy than their teacher is able to show?* An attempt has been made to analyze ability to estimate relative standing in United States geography. First, it was necessary to determine the relative order of ability in the subject. A criterion independent of teachers' marks or pupils' opinions was needed. In order to obtain this criterion a battery of seven United States geography tests was given to forty-four girls in a VIIB class. It was assumed that the tests employed actually measured ability in geography. The scores obtained in each test were ranked, and the ranks were averaged. Thus the final "Order of Ability in United States Geography" represented the average standing in seven tests. The ranking obtained served as a fair criterion for comparing judgments. The main purpose of the study was to compare the correlations between the three variables represented by:

1. Order of ability in United States geography.
2. The Teacher's estimates of ability in United States geography.
3. Pupils' estimates of classmates' abilities in United States geography.

After the data were gathered the following subsidiary problems presented themselves and may be stated as minor objectives of the study:

1. To determine the correlations between intelligence as measured by the National Intelligence Test, Scale A, Form 1, and
   (*a*) Scores in each of seven United States geography tests.
   (*b*) Average standing in seven United States geography tests.
   (*c*) Teacher's estimates of pupils' abilities.
   (*d*) Average ranking of pupils by pupils.
2. To determine the ability of each pupil in estimating the relative standing of classmates by correlating individual judgments with the ranking obtained by averaging results from seven objective tests.
3. To discover the standing in United States geography of the ten students whose judgments of classmates' abilities showed the highest correlations with the criterion (average rank in seven objective tests).
4. To study the tendency to over-estimate, or under-estimate, one's ability in the United States geography by comparing:
   (*a*) Each pupil's own estimate of her ability with the average of her classmates' estimates.
   (*b*) Each pupil's own estimate of her ability with her average standing in seven objective tests.

III. Gathering the Data

This experiment was conducted in a VIIB class in New York City.[1] The following tests were given over a period of two weeks to the forty-four subjects:

1. National Intelligence Test, Scale A, Form 1.
2. Courtis' "States and Important Cities of the United States." Form B.
3. The Gregory-Spencer Geography Tests, Form A (Grades VI, VII, VIII).
4. Posey-Van Wagenen Geography Scales, Thought R (Division 2), Grades VII, VIII.
5. Posey-Van Wagenen Geography Scales, Information R (General), (Division 2), Grades VII, VIII.
6. Buckingham and Stevenson's "Place Geography Test for United States."
7. Buckingham and Stevenson's "Information Problems," Form 1.
8. Thompson's "Minimum Essentials," Geography No. 6, North America.

The teacher's estimates of the pupils' abilities in United States Geography were obtained in three ways:

1. Geography marks for the month of April.
2. Ability estimated on a graphic rating scale.
3. An "Order of Merit" ranking extending from one to forty-four.

Each of these three sets of judgments made by the class teacher were reduced to rank order and averaged to obtain the teacher's estimate which was used as one of the factors of comparison in this study.

In order to secure a record of the pupils' estimates of classmates' abilities in United States geography the following technique was employed: An envelope containing forty-eight blank slips of paper was supplied to each girl. The teacher then dictated the names of the pupils. A single name was copied upon a slip. Next, the girls were instructed to sort out the names of the fifteen in the class who were "best" in United States geography. Then the names of the fifteen "poorest" students were sorted out. The remaining fourteen, presumably, represented those whose ability was considered average. The names of those in each of the three groups then were arranged in order of ability. At all times the raters were cautioned to keep in mind that they were considering "ability in United States geography." As a final step, all the slips were combined, and minor shifts in position

[1] This study was made possible through the cooperation of Professor Rudolf Pintner, Teachers College, Columbia University; Miss Ellen M. Phillips, principal of Public School 188G, Manhattan; Mrs. Anita H. Welle, assistant principal, and Miss Lillie Bohm, grade teacher. The following teachers in the school assisted—Miss Josephine Dalton, Miss Marie Hughes, and Miss Mary Schnepel.

TABLE I.—INTER-CORRELATIONS BETWEEN INTELLIGENCE TEST, UNITED STATES GEOGRAPGY TESTS AND ESTIMATES OF ABILITY MADE BY TEACHER AND FORTY-FOUR PUPILS

	National Intelligence Test	Courtis	Gregory-Spencer	Posey-Van Wagenen Thought R	Posey-Van Wagenen Information R	Buckingham and Stevenson's Place Geography	Buckingham and Stevenson's Information, Problem	Thompson's Minimum Essentials	Average rank in seven geography tests	April marks by class teacher	Graphic rating by teacher	Order of merit by teacher	Average of teacher's three estimates	Average ranking of pupils by pupils
National Intelligence Test		.17	.49	.49	.37	.39	.51	.42	.51	.21	.28	.24	.27	.40
Courtis	.17	...	.57	.09	.52	.58	.45	.70	.76	.20	.14	.09	.13	.02
Gregory-Spencer	.49	.57	...	.44	.65	.43	.57	.71	.83	.25	.28	.14	.24	.39
Posey-Van Wagenen Thought R	.49	.09	.44	...	.48	.22	.29	.37	.52	.29	.31	.20	.26	.36
Posey-Van Wagenen Information R	.37	.52	.65	.48	...	.54	.48	.61	.81	.38	.34	.26	.34	.31
Buckingham and Stevenson's Place Geography	.39	.58	.43	.22	.54	...	.17	.58	.68	.28	.24	.25	.27	.09
Buckingham and Stevenson's Information, Problem	.51	.45	.57	.29	.48	.17	...	.55	.68	.25	.21	.19	.21	.34
Thompson's Minimum Essentials	.42	.70	.71	.37	.61	.58	.55	...	.88	.37	.31	.24	.32	.34
Average rank in seven geography tests	.51	.76	.83	.52	.81	.68	.68	.88	...	.36	.32	.24	.35	.32
April marks by class teacher	.21	.20	.25	.29	.38	.28	.25	.37	.36	...	.88	.93	.96	.49
Graphic rating by teacher	.28	.14	.28	.31	.34	.24	.21	.31	.32	.88	...	.92	.96	.63
Order of merit by teacher	.24	.09	.14	.20	.26	.25	.19	.24	.24	.93	.92	...	.98	.47
Average of teacher's three estimates	.27	.13	.24	.26	.34	.27	.21	.32	.35	.96	.96	.98	...	.56
Average ranking of pupils by pupils	.40	.02	.39	.36	.31	.09	.34	.34	.32	.49	.63	.47	.56	

were made. The name slips then were numbered from one to forty-four; the papers returned to the envelope, and the envelope sealed. The name of the rater was written upon the outside of the envelope.

IV. Presentation of Results

The correlations between the fourteen variables considered in this study are presented in Table I. All were computed by means of Spearman's Rank Difference Method. Table II is presented to facilitate comparison of certain interesting relationships.

Table II.—Teacher's and Pupils' Abilities in Estimating Relative Standing in United States Geography (Forty-four Cases)

	r	PE
Teacher's estimate of pupils' geography ability and criterion (average ranking in seven United States geography tests)	+.35	±.09
Average ranking of pupils by pupils and criterion (average ranking in seven United States geography tests)	+.32	±.09
Teacher's estimate of pupils' geography ability and intelligence (N. I. T.)	+.27	±.09
Average ranking of pupils by pupils and intelligence (N. I. T.)	+.40	±.08
Intelligence and criterion (average ranking in seven United States geography tests)	+.51	±.08
Teacher's estimate of pupils' geography ability and average ranking of pupils by pupils	±.56	±.07

When we compute the correlation between the average ranking of pupils by pupils and the criterion, holding intelligence constant, the result is +.14. The correlation between the teacher's estimate of pupils' geography ability and the criterion, holding intelligence constant, is +.47.

The correlations between the separate geography tests and the National Intelligence Test ranged from +.17 in the case of the Courtis form to +.51 for the Buckingham and Stevenson Information Problems Test. The correlation between the intelligence test and the average ranking in seven United States geography tests was +.51.

The forty-four subjects varied greatly in their abilities in estimating the relative standing of one another in United States geography. The individual rankings were correlated with the criterion, made up of the average ranking in seven objective tests. One pupil's rankings correlated +.48 with the criterion; another correlated +.05 with the same measure. The median correlation for the forty-four subjects was +.27.

As was to be expected, there were some students whose rankings agreed with the criterion more closely than did those of their teacher.

It is interesting to note the class standing, the self-estimates, the classmates' estimates, and the intelligence of the ten students who were able to rank their classmates most accurately. Table III shows the correlations between these ten rankings and the criterion as well as the standing of these students on the basis of their own and their classmates' estimates.

TABLE III.—CORRELATIONS WITH CRITERION OF RANKINGS OF TEN STUDENTS AND CLASS STANDING OF EACH STUDENT IN A CLASS OF FORTY-FOUR

Student	A	B	C	D	E	F	G	H	I	J
Correlations with criterion (average seven tests)	.48	.47	.41	.41	.41	.40	.37	.36	.36	.36
Class standing on basis of ability	13	28	4	8	20	9	21	7	38	16
Rank assigned self	7	26	6	3	5	8	6	24	17	25
Rank assigned by classmates	20	17	14	22	7	40	29	27	26	29
Rank in intelligence	19	27	18	11	9	8	34	40	36	39

Eight out of ten of these "best" judges were in the upper half of the class, so far as their ability in United States geography was measured by means of seven objective tests. The rankings of Student I, who was thirty-eight in ability, were as accurate as those of her teacher. Student B, who was twenty-eighth in ability, apparently judged her classmates' geography ability more accurately than did her teacher. In intelligence, these "best" judges ranged from No. 8 to No. 40 in a class of forty-four. The geography ability of these ten students was estimated more accurately by themselves than by their classmates. There was just as much tendency for these ten pupils to over-rate themselves as for their classmates to under-rate them.

In this study, two criteria were available for measuring the tendency to over-estimate, or under-estimate ability in United States geography. When we utilize the criterion represented by the average ranking in seven objective tests, we find that twenty-seven of the students over-estimated their abilities on the average, 17.62 points. The seventeen who under-estimated their abilities, showed a negative displacement of 5.2 points. When we use the average ranking of pupils by pupils as the criterion, we find that forty out of forty-four girls over-estimated their abilities by 11.92 points. The four who ranked themselves lower than the average estimates of their classmates showed a negative disparity of only 3.09 points.

Somewhat similar observations have been made by Allport and Allport[1] who stated, "The possession of the trait in question seemed to insure a more accurate recognition of it, both in self and others, while those low in the desired qualities tended to over-rate themselves."

V. Summary and Conclusions

A group of forty-four VIIB girls in a New York City school served as subjects in an attempt to discover, in a particular class, whether or not the pooled judgments of the pupils correlated higher with a measure of ability in United States geography than did the ratings and grades given by their teacher. The members of the class were well acquainted. Most of them had lived in the same neighborhood and attended the same school for a number of years. They had been members of the same class for at least eight months previous to the time this study was made.

The members of the class were ranked from one to forty-four by each pupil composing the class. The teacher rated her students on a graphic rating scale; ranked them in order of merit, and also contributed the grades in geography for one month. A battery of seven United States geography tests was given. The average ranking in the seven objective tests served as the independent criterion against which was measured the accuracy of the judgments of the teacher and her pupils. The National Intelligence Test also was given.

The teacher's combined estimate of geography ability correlated $+.35 \pm .09$ with the composite criterion, made up of the average ranking in seven United States geography tests. The pupils' combined estimates of classmates' abilities correlated $+.32 \pm .09$ with the same criterion. The correlation between the teacher's estimate and the criterion, when intelligence was held constant, was $+.47$. The correlation between the pupils' estimates and the criterion dropped to $+.14$ when intelligence was held constant. It is evident that the opinion of this one teacher was of more value in estimating ability in United States geography than was the pooled estimate of the children composing the class. Apparently, the teacher was able to discriminate, at least to a higher degree than her pupils, between "ability in United States geography" and general intelligence. Whether or not this same condition exists, so far as other teachers and other classes are concerned, is not known.

[1] Allport, Floyd H. and Allport, Gordon W.: Personality Traits: Their Classification and Measurement. *The Journal of Abnormal Psychology and Social Psychology*, Vol. XVI, No. 1, April, 1921.

In the group of forty-four girls, there were ten whose ranking of their classmates correlated higher with the criterion than +.35. It will be remembered that +.35 was the correlation between the teacher's estimate and the composite criterion. Eight out of these ten students were in the upper half of the class in ability in United States geography. It does not follow, however, that the "best" in subject-matter are necessarily the "best" when it comes to estimating the relative abilities of classmates.

In intelligence, the ten "best" judges ranged from the eighth to the fortieth position in a group of forty-four. The geography ability of these ten students was estimated more accurately by themselves than by their classmates. There was just as much tendency for these ten pupils to overrate themselves as for their classmates to under-rate them.

When we utilized the criterion represented by the average ranking in seven objective tests, twenty-seven of the students over-estimated their abilities on the average, 17.62 points. The seventeen who under-estimated their abilities showed a negative displacement of 5.2 points. When we used the average ranking of pupils by pupils as the criterion, we found that forty of the forty-four girls over-estimated their abilities by 11.92 points. Four who under-estimated themselves, showed a negative disparity of 3.09 points. Depending upon our criterion, we may conclude that either sixty-one per cent or ninety per cent of our subjects over-estimated their abilities in geography. Logically, the former seems to be more nearly correct, since the extent of over-estimation is measured against a gauge of ability rather than against the average estimates of classmates. In the latter case the pooled opinion of the group emerges from the conflicting tendencies to over-estimate one's own ability and to under-estimate the abilities of one's associates.

This study has been productive of several results. A table showing the inter-correlations for seven widely used United States geography tests has been presented. Correlations between each test and a measure of general intelligence have been computed. Additional evidence has been obtained showing the extent to which children over-estimate their ability in school subject-matter. Finally, for at least one teacher and her class of forty-four girls, it has been shown that the teacher's estimate of the abilities of her pupils in United States geography is more accurate than the combined estimates of the children who compose the class.

EXPERIMENTAL COMPARISON OF TRUE-FALSE AND COMPLETION TESTS

VIOLET SHULSON

University of Idaho

AND

C. C. CRAWFORD

University of Southern California

Problem.—Two types of objective tests quite frequently used are the true-false and the completion types. It is the purpose of this article to compare these two types of tests to see which is the more valid measuring instrument and to discover any factors which might help in evaluating them for classroom use.

Procedure.—The procedure of the study was comparatively simple. Eight different experiments were performed, each involving a regular and an experimental class group. At the end of a period of study in a regular class, a test was given, consisting of twenty true-false items and twenty completion items. At the same time or later, this same test was given to students in an experimental class which had made no special study of the subject-matter covered by the test, to see what they could make on the test by guessing, bluffing, or using their general information. The principle underlying the whole procedure was that the better kind of test was that part of the total test which showed the more reliable difference in scores between the regular and the experimental classes. Thus, if the experimental class should be able to approximate the scores of the regular class more closely on the true-false than on the completion test, we should say that the true-false test was not as good a test as the completion.

The statistical method of comparing scores on the two kinds of classes was taken from McCall's "How to Experiment in Education." It consisted of computing the mean (M) for both the regular and experimental classes, subtracting these for the difference (D), and comparing the difference with its standard deviation (SDD) to determine its reliability. This resulted in the experimental coefficient (EC), which was computed from the formula $EC = \frac{D}{2.78SDD}$. An EC of 1.00 means that the difference is exactly 2.78 times its standard deviation, which is the same as saying that it is four times its probable error. Since an EC of 1.00 is considered "practical certainty," an

EC of 2.0 means that we are doubly certain that the difference is not due to chance errors of sampling. Furthermore, the type of test which, in a given experiment, yields the higher *EC* on comparing scores of the regular and experimental classes is the better measuring instrument, insofar as this investigation is able to determine values of the tests.

It should be pointed out here that it is entirely possible to compare the values of the two tests on other bases than the one used in this study. The fact that the completion test measures recall rather than recognition might justify its use even though inferior to the true-false test in statistical reliability. Likewise, the fact that the true-false test is scored more objectively might make up for any inferiority that might be found by the methods used in this study.

Classes Tested.—The students who were tested were college students in all but two experiments. In these two cases, the tests that were used for regular college classes were given to high school juniors and seniors, who served as experimental groups. In some cases, the regular classes consisted of college juniors and seniors, while the experimental classes were freshmen and sophomores. In other cases, the regular classes were freshmen and sophomores, and the experimental classes were juniors and seniors. In still other cases, there was no difference either way as regards classification.

Test Materials.—The tests used in the experiment were all of the home-made type, and were constructed by the teacher in charge of the regular classes as a part of the ordinary testing program.

Results.—Table I presents the results of the eight experiments in summarized form. Several items are to be noted in studying the table:

1. The mean score (*M*) for the true-false test averages higher than for the completion test throughout the eight experiments. The scores used here are the total number of right responses. If the right-minus-wrong formula were used, the results would be the reverse. The difference here is due to the greater possibility of guessing on the true-false test, which causes the scores to range between ten and twenty right, whereas the completion scores tend to range between zero and twenty right.

2. The standard deviations of the original scores (*SD*) on the completion tests are approximately twice those of the true-false tests. This is the natural result of the facts mentioned in (1) above, and suggests that when the two types are used side by side as a testing battery,

TABLE I.—SHOWING THE SUMMARY OF RESULTS OF EIGHT EXPERIMENTS DESIGNED TO COMPARE TRUE-FALSE AND COMPLETION TESTS

Experiment	Test	Cases (*N*)		Mean scores (*M*)		Standard deviations of scores (*SD*)		Standard deviations of mean scores (*SDM*)		Differences between mean scores (*D*)	Standard deviations of the differences (*SDD*)	Experimental coefficients (*EC*)	Tests favored
		Regular	Experimental	Regular	Experimental	Regular	Experimental	Regular	Experimental				
1	Completion	30	22	10.40	4.54	3.08	1.89	.56	.40	5.86	.68	3.1	Completion
	True-false	30	22	16.27	11.82	2.20	1.82	.40	.39	4.45	.56	2.9	
2	Completion	16	16	10.75	6.00	3.60	2.09	.90	.52	4.75	1.04	1.6	True-false
	True-false	16	16	16.31	12.00	1.79	2.03	.45	.51	4.31	.68	2.3	
3	Completion	137	20	13.52	7.35	3.31	2.46	.28	.56	6.17	.62	3.6	Completion
	True-false	137	20	17.02	16.55	1.57	1.12	.13	.25	.47	.28	.6	
4	Completion	31	27	11.16	6.11	3.96	3.33	.71	.64	5.06	.95	1.9	True-false
	True-false	31	27	15.26	11.44	2.12	2.38	.38	.46	3.82	.59	2.3	
5	Completion	36	25	11.75	6.08	4.00	3.50	.67	.70	5.67	.96	2.1	True-false
	True-false	36	25	16.64	12.84	1.92	1.93	.32	.39	3.80	.50	2.6	
6	Completion	137	26	13.52	7.62	3.20	3.50	.27	.69	5.90	.74	2.9	Completion
	True-false	137	26	17.02	15.81	1.57	1.96	.13	.17	1.21	.22	2.0	
7	Completion	36	14	11.75	3.21	4.01	2.01	.67	.54	8.54	.86	3.6	Completion
	True-false	36	14	16.64	11.00	1.92	2.25	.32	.60	5.64	.68	3.0	
8	Completion	36	35	11.75	3.40	4.01	1.94	.67	.33	8.35	.75	4.0	True-false
	True-false	36	35	16.64	11.71	1.92	1.54	.32	.26	4.93	.43	4.1	

NOTE.—The table should be read as follows: "In the case of Experiment 1, the completion test was taken by thirty students in a regular class, and by twenty-two students in an experimental class; the mean score in the regular class was 10.40 and in the experimental class, 4.54, etc.

the true-false scores should be multiplied by two before being combined with the completion scores. Otherwise, due to the greater variability of the completion scores, they would have approximately twice the weight of the true-false scores in determining the composite score.

3. The third fact to note is that in every experiment the completion test gives a larger numerical difference between regular and experimental classes than the true-false test, when measured in terms of raw scores. This is due simply to the greater variability of the completion scores, and has very little significance at all for our study, except that it is likely to be misinterpreted unless examined critically.

4. The most important outcome in the whole experiment is to be found in the *EC* column of the table. We see that out of the total eight experiments, four were favorable to the true-false and four to the completion test. In other words, according to the statistical procedure used in this study, the two tests are equally good, and equally able to distinguish the students who have studied the lesson from those who have not.

Certain facts were revealed by the experiments that are not shown by the figures in the table and which are, nevertheless, important. One is that the scoring of the completion test is far from objective and constitutes a source of embarrassment because of the problem of synonym and near-synonym answers. One test was scored by two different persons with the result that in some papers four or five questions out of twenty were given opposite grades. Another non-statistical fact to be noted is that the completion test measures recall somewhat more than the true-false test and is, therefore, a valuable supplement to it, even though not necessarily any better.

Conclusion.—The conclusion, therefore, seems to be that, while the completion test lacks objectivity, it is as valid as the true-false test and that the two should be used together to supplement each other by measuring different types of mental reactions.

THEORIES UNDERLYING THE STATISTICAL DETERMINATION OF CREDIT TO BE ALLOTTED ITEM RESPONSES

ELISHA F. BLISS, JR.

Hopewell, N. J.

There has been some confusion as to the right statistical procedure for determining the "correctness" of test item responses, due primarily to the fact that the underlying theories have not been well defined. It is the purpose of this paper to call attention to some of these theories in an attempt to clear up this situation.

An item is here defined as a single task; a test is a collection of items. By criterion scores is meant the best estimate obtainable of an individual's ability in the trait in question. It is assumed that the response which an individual makes to an item determines the credit, or score, which he will obtain upon the item.

Since it is possible to correlate item scores against criterion scores, it holds that the amount of credit to be given any response is a function of the mean of the criterion scores of the individuals giving that response. We find, therefore that, given the means of the criterion scores of various responses to an item, we can arrange these responses in the order of the amount of credit to be accorded to them; although, until we make assumptions concerning the sigma of the item scores, we cannot determine the amount of credit.

Since the mean criterion scores of the various responses to an item are subject to considerable error unless a very large number of cases is used in analyzing the item, the hypothesis has been advanced that the relative frequency with which a response occurs is an index of the amount of credit which should be accorded that response.

From the preceding discussion it would appear impossible to discover the amount of credit, or order of credit from the frequency alone. In practice, we rarely, if ever, come upon an item whose responses are sufficiently unselected to permit a theoretically ideal distribution of scores; we find the differences in score value between successive responses are not equal. Furthermore, while in unlimited response material there is no lower limit to the score which an individual may make on an item, there is usually a higher limit in what is logically the "correct" response. The effect of this higher limit is that in the frequency with which the "correct" response occurs is included all

those cases which would exceed the score of this response if the higher limit of score on the item were extended. Being the sum of the frequencies of several responses instead of the frequency of a single response, it is often found to exceed the frequency of any other response. The observation of this fact has led to the erroneous conclusion that the response occurring most frequently is necessarily the "correct" response.

From this it will be seen that the frequency with which a response occurs is of value in checking the "correctness" of a response. When the most frequent response is not the one deemed "correct," it is well to compare the mean criterion scores of the two responses; in cases where no criterion is available, a subjective examination of the item will confirm the "correct" response, expose an error in the original judgment, or else show the item to be definitely ambiguous.

Due to the possible unequal differences between score values of the responses to an item, there appears to be little ground for the belief that other responses than the correct one should receive credit either in the order of, or in proportion to the frequency with which they occur.

That crediting the most frequent response as "correct" will yield the greatest validity (correlation between criterion and item scores) for an item will appear false except in instances when the frequency of the better responses may be so small that chance may play an important factor in determining the mean criterion scores of those responses.

THE SPEARMAN-BROWN PROPHECY FORMULA AND THE SEASHORE TESTS

PAUL R. FARNSWORTH

Stanford University

Wherever reliabilities of tests are being ascertained, one finds in general use what is popularly called the Spearman-Brown prophecy formula. One of its functions seems to be that of prophesying the correlation between two forms of a test, (x and y), from the intercorrelations of two comparable halves, (1 and 2) of one of the forms. The formula may be stated $r \times y = \frac{2r^{12}}{1 + r_{12}}$. Ruch, Ackerman and Jackson[1] and others have demonstrated its approximate validity with educational material.

Lanier,[2] however, has attempted to show that the formula does not function for the Seashore battery of sound tests. He did this by stepping up the odd-even and top-bottom-half intercorrelations of these tests and finding that the augmented correlations were not comparable to the retest correlations. In the writer's judgment this is not always a wise procedure. Retests are frequently not comparable to two forms of a test. This should be especially obvious with the Seashore tests. These tests are long and extremely monotonous. Although one can usually keep up the interest for one rendition, repetition seems to be boring. Even when the tester informs the subjects that he is giving them new forms, it is soon apparent to many that the same uninteresting variety of test is at hand.

Other modes of approach to this problem of checking the Spearman-Brown Formula are possible.[3] Several members of the Seashore battery contain one hundred items each.[4] This presumably makes them sufficiently long to be broken into two forms. Such a treatment is especially easy in the intensity test. Face A of the record is intended

[1] Ruch, G. M., Ackerman, L. and Jackson, J. D.: An Empirical Study of the Spearman-Brown Formula as Applied to Educational Material. *Journal Educational Psychology*, Vol. XVII, 1926, pp. 309–313.

[2] Lanier, L. H.: Reliability of Mental Tests. *Journal of Experimental Psychology*, Vol. X, 1927, pp. 69–114.

[3] In the present article, students of Stanford University and the State University of Iowa served as subjects. The writer is indebteded to Dr. Robert H. Seashore for access to the Iowa data.

[4] Seashore, C. E.: "Manual of Directions and Interpretations, Etc." Columbia Record Co.

to be the equivalent of face B. The two faces, therefore, may be considered as two forms of the same test.

TABLE I.—INTENSITY TEST (N = 328)

	r	PE	Raised by Spearman-Brown
r between faces A and B	.507	.027	
r between odds and evens, face A	.289	.034	.448
r between odds and evens, face B	.392	.031	.563

No other test in the Seashore battery is arranged in this manner. The "time" test, however, may be split into two forms of alternate columns. That is, one form can be made from columns A, C, E, G and I, and the other from the remaining five columns B, D, F, H and J.

TABLE II.—TIME TEST (N = 142)

	r	PE	Raised by Spearman-Brown
r between alternate columns	.520	.042	
r between odds and evens (A, C, E, G, I)	.374	.049	.544
r between odds and evens (B, D, F, H, J)	.382	.049	.553

The tests of "pitch," "intensity," and "time" may be treated in still a different manner. According to the approved method of procedure, the responses on these tests are tabulated into ten columns and ten rows. The items of each column are intended to be of equivalent difficulty. It is thus possible to divide the tests into approximately equal halves by cutting off the upper five rows from the lower five.

TABLE III

	r	PE	Raised by Spearman-Brown
Pitch (N = 328)			
r between upper five and lower five rows	.724	.018	
r between odds and evens, upper five rows	.573	.025	.729
r between odds and evens, lower five rows	.616	.023	.762
Intensity (N = 328)			
r between upper five and lower five rows	.508	.028	
r between odds and evens, upper five rows	.390	.032	.561
r between odds and evens, lower five rows	.440	.030	.611
Intensity (N = 216)			
r between upper five and lower five rows	.659	.026	
r between odds and evens, upper five rows	.459	.036	.629
r between odds and evens, lower five rows	.510	.034	.675
Time (N = 142)			
r between upper five and lower five rows	.460	.044	
r between odds and evens, upper five rows	.367	.049	.537
r between odds and evens, lower five rows	.307	.051	.470

SUMMARY

On a study of the Seashore battery of sound tests, Lanier had previously found that the Spearman-Brown prophecy formula failed to function when checked against retests. In this study two forms of each test were made from a single rendition. The formula was found to predict with considerable accuracy in most instances.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY JAMES E. MENDENHALL
Research Associate, Institute of Educational Research, Teachers College
Columbia University

Educational Psychology

General Review and Summary: Educational Psychology. V. A. C. Henmon. Psychological Bulletin, July, 1928, pp. 377–388.

The Testing Movement in the Light of Recent Research. Edward L. Thorndike. Journal of Educational Research, May, 1928, pp. 345–349.

Concerning Subjective Judgment of Difficulty. Eric Farmer. The British Journal of Psychology, April, 1928, pp. 438–442. "Those who differ most from the normal are the worst judges of their own performance (in ability to make hand directed pointer coincide with mechanically driven pointer-Pursuit Meter)."

Child Psychology

Children's Activities and the School Curriculum. J. L. Meriam. School and Society, April 14, 1928, 458–462. The study is based upon the observation of one hundred fifty children between ages six and twelve. The frequency of activities is used to determine the major ways in which children spend their time.

Sex Differences in Reference to Reading Books, Just for Fun. Harvey C. Lehman and Paul A. Witty. Education, June, 1928, 602–617. Girls exceeded boys in ability at the age of fourteen and interest in reading between ages 10½ and 16½.

The Influence of Chronological Age Versus Mental Age on Play Behavior. Harvey C. Lehman and Doxey A. Wilkerson. Journal of Genetic Psychology, June, 1928, 312–324. Data regarding play behavior of six thousand children was utilized. "Chronological age appeared to be more potent than mental age in influencing the boys 'play behavior.'"

Some Compensatory Mechanisms of the Negro. Harvey C. Lehman and Paul A. Witty. The Journal of Abnormal and Social Psychology, April–June, 1928, 28–37. Negro children engage considerably more frequently than white children in the following play activities: church activities, playing school, writing poetry, social activities, boxing.

The Sleep of Young Children. Josephine C. Foster, Florence L. Goodenough, John E. Anderson. Journal of Genetic Psychology, June, 1928, 201–217. Questionnaires regarding the amount of sleep of one thousand one hundred eighty-six children under eight years of age were collected.

Administration and Supervision

A Study of High School Seniors to Determine Who Shall be Recommended to College. Celia D. Pierson and Charles H. Nettels. School and Society, August 18,

1928, 215–216. The regression equation by which to predict college scholarship freshman year equals .23I (Terman Group Test Score) +.96H (High School Marks) +.047C (Character Ratings) −29.97.

Predicting College Success for the High School Senior. J. B. Johnston. The Vocational Guidance Magazine, April, 1928, 289–294.

The Student Who Drops Out of College. George R. Moon. School and Society, May 12, 1928, 576–578. A University of Chicago study of replies received from two hundred thirty-nine dropouts (after freshman year) indicates these four chief reasons for leaving (aside from scholarship): Finances, health, home conditions, and some form of dissatisfaction.

The Effect of Class Size on Scholastic Acquirement in Educational Psychology. B. F. Holland. School and Society, June 2, 1928, 668–670. Four hundred ninety-two university students were tested. The writer concludes; "that students taught by the lecture method acquired no less of the course subject-matter when grouped in large classes than when grouped in small classes . . . Such variables as the day of the week and hour of day and the presence of upper classmen in freshman classes had no appreciable effect on the average achievement scores." Correlations between achievement and intelligence of .48, .59, and .70 are reported.

Do Teachers Improve with Experience? J. E. Bathurst. The Personnel Journal, June, 1928, 54–57. "Age and experience of one hundred seventy-one high school teachers were compared with scores in the Aptitude Tests for High School Teachers. The zero order coefficient of correlation (1) between experience and score was 0.15, (2) between age and score, 0.06; and (3) between age and experience, 0.88. Teacher efficiency decreases slightly with age if experience is kept constant, and increases slightly with experience if age is kept constant.

University Records of Students from Junior Colleges. Walter Crosby Fells. California Quarterly of Secondary Education, June, 1928, 301–317. At Stanford University, the junior college graduate entrants are superior to native Stanford students in the Thorndike Intelligence test and in academic achievement.

Personal Record Cards for Schools and Colleges. Ben D. Wood. The Educational Record, July, 1928, 14–52. Suggested cumulative educational record forms.

Failures and Double Promotions in the Elementary School. H. A. Greene and G. Freden. Journal of Educational Research, April, 1928, 262–272. A study of 37,330 Iowa school children in Grades I to VIII.

An Evaluation of Items to Observe in Classroom Supervision. A. L. Barr. Journal of Educational Research, June, 1928, 53–65. An attempt is made to analyze teaching practices and to validate observations of them.

Educational and Vocational Guidance

Personnel Factors and Turnover of Sales Clerks. C. J. Ho. Industrial Psychology, August, 1928, 358–361. The associations of sex, age, marital status, education, and intelligence with turnover are presented.

Extra-curricular Activity and Success. R. E. Thornhill and Carney Landis. School and Society, July 28, 1928, 117–120. A study of the extra-curricular activities and of the academic records of Who's Who men from Wesleyan University indicates that this group engaged in more journalistic, more dramatic and more scholastic activities, less athletic, and less political activities than did the index group.

Age of Graduation and the Ph.D. E. Earle Franklin. School and Society, August 11, 1928, 182–184. A report concerning one thousand three hundred eighty-one men and women who secured doctor's degrees at John-Hopkins during a period of fifty years.

Does It Pay to Change Jobs? Harry D. Kitson and Noel Keys. The Personnel Journal, June, 1928, 33–37. "A study was made of the vocational histories of one hundred seventy high grade male clerical workers, twenty to thirty-one years of age inclusive . . . First computation showed a low correlation of .23 between length of service record and earning power; with age constant of .04 . . . Increase in earning power is likely to depend on such factors as intelligence, initiative, local opportunity and the like, rather than on mere time-serving."

Newer Personnel Practices in Colleges. William F. Book. The Personnel Journal, June, 1928, 38–53. A brief report of significant aspects of this work at Indiana University.

The Prediction from Photographs of Success and Vocational Aptitude. Carney Landis and L. W. Phelps. Journal of Experimental Psychology, August, 1928, 313–324. On the basis of photographs the judges (college students) were unable to distinguish between successful and unsuccessful men.

Measurement of Special Abilities

General Review and Summary: Tests and Measurements in Music. Jacob Kwalwasser. Psychological Bulletin, May, 1928, 284–301.

The Role of Consonance and Dissonance in Music. Martha Guernsey. The American Journal of Psychology, April, 1928, 173–204. A study of the æsthetic judgments of one hundred six subjects (trained and untrained) regarding various tonal combinations.

The Affective Characters of the Major and Minor Modes in Music. Christian Paul Heinlein. The Journal of Comparative Psychology, April, 1928, 101–142. A study of ability to discriminate chords and of the feeling tones evoked by certain chords. Thirty college subjects, trained and untrained, constituted the experimental group.

A Test for Motor Capacities in the Industries and in the School. Richard L. Schultz. The Journal of Applied Psychology, April, 1928, 169–189. Motor capacity (as measured by the Viteles Machine Feeding Test is investigated relative to intelligence, sex, chronological age, race, achievement in school subjects of three hundred two cases (CA11 to 15).

A New Method of Analyzing Muscial Style by Means of the Reproducing Piano. Guy M. Whipple. The Journal of Applied Psychology, April, 1928, 200–213. By means of the player roll, the notes played by master pianists are analyzed into "intensity" and "duration."

The Motor Control of Children as Involved in the Dressing Process. Loirsa C. Wagoner and Edna M. Armstrong. Journal of Genetic Psychology, March, 1928, 84–97. The authors present the correlations between a buttoning test, Stanford-Binet, Merrill-Palmer, Goodenough. Drawing based upon thirty children CA1–9 to 5–1.

A Measurement of Mechanical Interests. Ruth M. Hubbard. Journal of Genetic Psychology, June, 1928, 229–254. The items of an interest analysis questionnaire were validated by a "shop ability" test and by mechanical ability

tests. Interest analysis and shop criterion are correlated with other variables—socio-economic ratings, questionnaires, and physical ability tests.

The Organization of Mechanical Intelligence. Harold Dean Carter. Journal of Genetic Psychology, June, 1928, 270–285. Inter-correlations ($N = 100$) between Shop Operations, Minnesota Assembly, Spatial Relations, Paper Form-board, Information Criterion, Interest Analysis, Son's Mechanical Operations, Stenquist Picture II, etc. are presented. Correlations of columns ranged from $-.61$ to $+.93$, average $+.30$. The writer concludes "that there are factors of mental ability which are quasi-general, running through about four or five tests only," and that there are group factors that overlap.

Study Habits and Methods

The How to Study Problem. Joseph S. Butterworth. Journal of Educational Research, June, 1928, 66–76. The study habits of ninety-eight teachers college freshmen are examined for significant differences in method of study and in method effectiveness.

Miscellaneous

The Case Problems of Student-teachers. H. A. Sprague. Educational Administration and Supervision, May, 1928, 314–324. A method for training students by the consideration of problem cases is suggested.

A Problem in Social Adjustment. Anna Spiesman Starr. The Psychological Clinic, May-June, 1928, 85–96. A statistical study of 18,395 cases examined at the Neuro-psychiatric Clinic of the Municipal Court of Philadelphia. Conclusions; 23.7 per cent were diagnosed as normal; 36.8 per cent as sub-normal; 19.3, as psychoneurotic, etc.

Value of a College General-science Course. Edward S. Jones. School and Society, May 19, 1928, 607–608. A report of the preferences of one hundred twenty-five students regarding a general physical science course, a general social science course, and other college courses.

Student-teachers and the Thurstone Psychological Examination. School and Society, May 26, 1928, 638–640. At the University of California two hundred seventy-six student-teachers (fourth year students) made scores significantly higher than do college freshman according to the norms.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

Creative Music

Creative Music in the Home, by Satis N. Coleman. Valparaiso, Indiana: Lewis E. Myers and Co., 1928. Pp. 400.

The Psaltery Book, by Satis N. Coleman. New York: Creative Music Studio, 15 Claremont Ave., 1928. Pp. 78.

Bells, by Satis N. Coleman. New York: Rand McNally and Co., 1928. Pp. 462.

It is scarcely fair to Mrs. Coleman's scholarship to try to review in a few hundred words more than ten years of experiment in the field of music for children. Her heresies and the studies conducted to justify them are both too complex to admit of superficial treatment. However, these last three books reveal clearly the theory underlying the author's work and the results which might be obtained. "Creative Music in the Home" and "The Psaltery Book" illustrate the principle that children should "make" music; "Bells," the principle that they should "know about" music.

These two ideas form the nucleus of Mrs. Coleman's theory. In an earlier work[1] she says," We shall study and make all types of instruments . . . They (the children) shall find out how the art of music was evolved and by their own creative work they shall experience its most important stages." And again, "If a child *lives* the art of music from its primitive beginnings, makes his own instruments and plays on them, and discovers for himself each stage in the development of musical instruments, how can he help being musical?"

With these views, then, her task was "to bring the whole history of music within reach of the child," and to plan a course which would "give children the actual experience of developing a complete art of music." A tremendous undertaking, but one which places Mrs. Coleman as one of the few scholars in the field of music education.

[1] "Creative Music for Children." G. B. Putnam's Sons, New York, 1922.

Even a superficial glance at the books will convince the reader of the amount of careful research and experiment which they entailed.

"Creative Music in the Home" shows the development of instruments from the very simplest ones, which can be made from a variety of crude materials, such as cocoanuts, nail-kegs, and candy boxes, to those of the modern symphony orchestra. Much historical material about these instruments and their uses is interwoven with the descriptions of how they are made. The illustrations are abundant and so placed that they supplement the reading and make it meaningful. "The Psaltery Book" tells how to learn to play the psaltery. Mrs. Coleman thinks that this sweet-toned and restful instrument, as she calls it, may have a real place in modern life. "Bells" is a very inclusive history of apparently all the bells in the world, both ancient and modern. Many fascinating and profitable units of work could be developed from its use in the upper grades. As a matter of fact, the book itself grew out of an interest in bells in one of the classes at Lincoln School, New York City, where Mrs. Coleman has done much of her work. It must also be said here that the program of school music comprises more than the making of instruments and reading about them. There is much improvisation of melodies, singing, and ensemble playing.

In fact, the plan which has developed during these years of study seems almost ideal if one accepts Mrs. Coleman's original hypotheses. But herein lies the difficulty. Is it safe to assume that children cannot help being musical, given the experiences which she suggests? Is the making of instruments intrinsic to music experience? Or, is it not a more or less overt expression of manipulation rather than one of feeling and imagination? Does the procedure provide for all-round growth in both appreciation of and participation in music?

In my judgment these are not mere academic questions. They present some of the fundamental issues with which students of this problem are confronted. For example, another large group, believes that the first steps in the program of music are "listening, thinking, and then playing." Philosophically, both views are legitimate. But what is needed now is not so much philosophical discussion as study and research to learn the psychology of both. Such experiment may reveal that Mrs. Coleman has sacrificed the *what* of music for the *why* and *how*. It seems to me that in teaching music, as well as literature, one should surround the child with the very best that has been written—at his level, of course—in addition to stimulating crea-

tive effort. However, that is also personal opinion. It remains for a student with a truly scientific attitude to provide the evidence. Now that the content of Mrs. Coleman's course is fairly complete, she herself may wish to know more of its psychology.

LOUISE KRUEGER.
Assistant, Teachers College, Columbia University.

HOW THE SAMOAN FACES ADOLESCENCE

Coming of Age in Samoa, by Margaret Mead. New York: William Morrow and Co., 1928, Pp. XV + 297. $3.00.

An American public, acquainted with Samoa largely through motion pictures made on the California coast, may well be expected to view with alarm the invasion by an anthropologist of this geographical area hitherto so peculiarly set aside as the exclusive domain of the writer of fiction and the director of films that are stranger than fiction. All too many of that public, should Miss Mead's "Coming of Age in Samoa" fall into their hands, will never enjoy the one-hundred eighty pages of delightful description of Samoan village life. They will lay the book aside without passing beyond the brief introductory statement of the serious scientific purpose of the study, persuaded that no one entertaining so serious an intent could write in so absorbingly interesting a style as Miss Mead achieves nor, so without the injection into the narrative of moralizing comment on the contrasts to American civilization presented by the light hearted casualness of Samoan life.

That they should be deterred from enjoying the narrative is unfortunate, for having read it, the momentum of its charm might easily carry them through the remaining two chapters with their keen analysis of the problems of the adolescent, problems which Miss Mead says, in her introduction, "have terrified the cautious reactionaries, tempted the radical propagandist to missionary crusades among the defenceless youth, and worried the least thoughtful among us."

To discover whether the apparent difficulties of being adolescent are really such, or are only difficulties of being adolescent in America was Miss Mead's purpose in making a detailed study of a group of adolescent Samoan girls. Against the background of the simple pattern of Samoan life is thrown into contrast the complexity of the opposing forces operating to produce disturbing stresses in youth

under the conditions of modern civilization. An analysis is made of the perplexing choices which must be made; of the absence of adequate guidance in making such choices; and of the lack of satisfactory finality in what seem to be crucial elections. So pertinent and convincing are Miss Mead's comments upon the adequacy of the assistance furnished by society for young people confronted by problems multiplied beyond their ability to cope with them, that the reader interested in social problems wishes that she might write at greater length upon education. In this day of multiplied educational writings it is no mean accomplishment to have written a book which leaves the impression that the author has told far from all that he knows in the field in which he writes.

LESTER M. WILSON.
International Institute, Teachers College.

EXTRA CURRICULAR ACTIVITIES

Extra Classroom Activities, by Riverda H. Jordan. New York: Thomas Y. Crowell Co., 1928. Pp. X + 302. $2.50.

This text appears in a field where numerous volumes, chiefly devoted to junior and senior high schools, have been published in the past half-dozen years. Believing in the "essential unity of the entire movement" from the elementary school through the senior high school, this book justifies itself by undertaking to discuss extra-classroom activities for the elementary as well as the junior and senior high schools. It is largely a survey of existing forms of such activities with running comments upon difficulties to be avoided—such comments as might be garnered from alert principals over a long period of years. As such, it may be useful to either present or embryo administrators who have not had special training in the field and have no time to be over-thorough.

To a more serious student the text is lacking in a fundamental and comprehensive philosophy of the purposes and possibilities of extra-curricular activities. Historically, these activities have grown up much as Topsy did. Recognition of this character of their growth has resulted in some thorough-going efforts to ascertain causes for their growth and first principles for their control and guidance, and their relationships to the curricular activities of the school. A principal, alert and sensitive to possibilities in this field seeking to orient himself with regard to major principles and methods of procedure,

would find this inadequate answer in the text: Spontaneous activity of a social (or sociable—the two terms are interchangeable here) nature with as little supervision as possible! Seemingly, the widespread practice of devoting definite time in the school day to extra-curricular activities and the fundamentals of building up such a program through both teachers and pupils is ignored. Occasionally "character building" (that time-honored excuse of last resort since the founding of denominational colleges for any activity in the curriculum) is mentioned as a guiding principle. No measuring stick is offered by which the uninitiated may judge spontaneity or good character. The text does not point out with any clearness or enthusiasm the possibilities in student organizations for training in the administration of a democracy.

Most of the treatment is too extensive. It lacks thoroughness. But six pages are devoted to elementary school athletics! The elementary school is inadequately handled throughout. There is no discussion, for example, of psychological reasons for different kinds of hobby clubs and student organizations in the lower schools than obtain in the higher schools. Such a discussion is sorely needed. Only one type of debate-judging (in high school) is mentioned. Music courses, in which the student produces music, are recommended as constants for all junior and senior high school pupils in spite of the fact that most of us are only consumers and might become consumers of good music if given the opportunity to learn what is good. It is interesting to note here that music was considered as an extra-classroom activity originally in the text. Making some form of it a constant removes it from that field to the curriculum. Some writers hold that one of the purposes of extra-curricular activities is to make such contributions to the curriculum. As a principle, one searches vainly for a statement or defense of it within the text.

The book has a pleasing appearance and is printed to be read easily. Problems at the end of each chapter make it adaptable to classroom teaching. The bibliography is excellent and an appendix contains a number of constitutions for student body organizations. The chapter on handling the finances of various school societies and organizations is well done. In the chapter on music there is a good explanation of the manner in which one school orchestra was developed. The method was fundamental, and applied to the development of practically any extra-curricular activity. It warranted intensive study because principles of method are paramount in the establishment of worthwhile extra-curricular activity programs.

The book fills a need temporarily but will surely be superceded by one of greater thoroughness giving more attention to basic principles.

J. H. Coleman.

Graduate Student Teachers College.

Newer Methods of Teaching

Inspirational Teaching. A Record of Experimental Work in the Teaching of English, by George Mackaness. New York, E. P. Dutton and Co., 1928. Pp. 225.

"Inspirational" anything is a red rag to those of us whose enthusiasms are not unmixed with common sense. Mr. Mackaness's book deserves a better title. It is no pseudosermon full of vague uplift. It is provocative, imaginative, amusingly written, packed solidly with hundreds of novel helps and suggestions in methods of teaching literature for enjoyment and understanding. It has excellent reading-lists for youngsters and good, though brief bibliographical material for teachers. A sort of handbook of newer method, if you like, replete with examples of how those methods worked or failed with the pupils. Written for teachers of English in the Australian schools, it is well abreast of modern practice everywhere, quite au courant with the field of drama, poetry, fiction and belles lettres in America and Europe. "Inspirational Teaching" believes in the power of the teacher; but it derives its impetus from a much more important conviction. The secret of effective teaching, to Mr. Mackaness, lies in suggestion and invitation. He is very clear on the point that the *pupil* is an almost untapped reservoir of creative activity and appreciation, and he shows over and over again that there are no limits, practically, to what one may draw forth, provided one is careful not to pour too much in. In America, Dewey and Kilpatrick bear witness to the philosophical integrity of the method. In England and Australia, however, missionary work would seem more necessary, from the author's unstilted insistence on the point. He seems to dread a little that his idea may sound like a crackbrained plea for professional laziness.

The present reviewer is in hearty accord with all but two of Mr. Mackaness's theories and practices. His pupil graphs, his delight in the play-way, his "magimaps" (a marvellous device for loosing a whole flood of imaginative, creative writing), his "lecturettes," his methods of correcting errors, seem very sound, all of them, and are all most

appealing. Certainly he wishes Mr. Mackaness were teaching next door to him, a focus of thrilling activity in English literature and idiom. But despite Australia's Q. C. (Qualifying Certificate Examination) and our own College Entrance Examination, it is hard to subscribe to the following as vita[illegible]ugh to be a "first duty:"

> Thus your first duty will be to organize a course in formal grammar and in parsing and analysis as a sure and certain basis before proceeding to the more interesting and imaginative sections of English teaching.

Nor, with the work of Hughes Mearns in mind, is one ready to agree that the way to teach children to write poetry is from without, by mere parody, imitation of models, and sketchy instruction in the use of rhymes and meters.

B. J. Stolper.
The Lincoln School, Teachers College.

A Textbook of Individual Differences in Native Traits

The Psychology of Individual Differences, by Robert S. Ellis. New York: D. Appleton and Co., 1928. Pp. XXIII + 533.

Contributions to differential psychology have appeared in psychological journals with increasing frequency for many years. Courses in the psychology of individual differences are also on the increase both in regular and summer sessions. The general practice in such courses has been to refer students to Thorndike's third volume or his "Individuality," perhaps to Woodrow's "Brightness and Dullness in Children," and to many select references in psychological journals. There is need for a book usable as a text which covers this aspect of psychology. In this volume Dr. Ellis not only attempts to supply such materials of instruction, but also endeavors to make it intelligible to students of average college ability. As a consequence the first part of the book is largely taken up with the consideration of background materials. This includes brief consideration of such topics as scientific procedure; statistical procedure and interpretation; relevant sections from neurology, physiology and endocrinology; and a lengthier treatment of the facts of inheritance.

So much for background. The subject-matter proper may be classified under three large captions: (1) Centributing factors of individual differences—including sex, race, age, environment. (2) Extreme deviations in mental traits—including genius, mental deficiency, insanity, delinquency and criminality. (3) Applications and implications—including a chapter each on Social Differentiation, The

Organization of Mind, Applications to Education, Applications to Business, Applications to Law and Government, and Miscellaneous Applications. This last includes individual hygiene, individual differences in religion, marriage and divorce and interpretation of history. Surely, enough material for a course for students of average college ability without adequate psychological background! That is what the author aimed the book to be. It foots this bill.

But it does not very satisfactorily do what the publishers say it does, namely, answer the question "Why do I behave like myself and no one else?" The more usual meaning of such a question implies a psychiatric approach and an interpretation in which motivation of behavior is stressed. In this book, however, as the author informs us in the preface, "The point of view is essentially biological, experimental, and statistical and but little attention has been given to pathology and psychoanalysis. Individual differences in acquired traits are discussed to some extent, but the field primarily considered is differences in native traits. Differences in acquired traits are held to be of interest chiefly to the educational psychologist and the social psychologist." From a field of investigation so limited with no mention of Hull's findings concerning the nature of trait differences in every individual as well as individual differences in every trait—which is based on non-psychoanalytic findings—it would be folly to expect a revealing of "why each of us behaves like himself and like no one else." The answer is not here. But for the purpose the author had in mind, the book is well-planned and fairly adequate. H. MELTZER.

Psychiatric Clinic. St. Louis.